

AD _____

Award Number: DAMD17-98-1-8045

TITLE: Improving Clinical Diagnosis Through Change Detection in Mammography

PRINCIPAL INVESTIGATOR: Yue-Joseph Wang, Ph.D.

CONTRACTING ORGANIZATION: The Catholic University of America
Washington, DC 20064

REPORT DATE: September 2001

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20020124 252

REPORT DOCUMENTATION PAGEForm Approved
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE September 2001	3. REPORT TYPE AND DATES COVERED Annual Summary (1 Sep 00 - 31 Aug 01)	
4. TITLE AND SUBTITLE Improving Clinical Diagnosis Through Change Detection in Mammography			5. FUNDING NUMBERS DAMD17-98-1-8045	
6. AUTHOR(S) Yue-Joseph Wang, Ph.D.				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The Catholic University of America Washington, DC 20064 E-Mail: wang@pluto.ee.cua.edu			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited				12b. DISTRIBUTION CODE
13. ABSTRACT (Maximum 200 Words) <p>Temporal change of mass lesions overtime is a key piece of information in computer-aided diagnosis of breast cancer and treatment monitoring. For a specific patient, change detection is a critical step to select lesion candidates for follow-up diagnosis performed by either clinicians or computers. In the third year of this project, we developed a hybrid image registration technique to align temporal sequences of the same patient and detect changing lesions, and developed a neural network based classifier to derive the probabilities of true masses.</p> <p>In particular, we developed (1) mPAR and MLP based registration algorithm to recover non-rigid deformation; (2) a new change detection scheme using independent component analysis of image sequences; (3) a feature extraction algorithm to obtain discriminative imagery features of true masses against mass-like normal tissues; and (4) a neural network based decision support system for mass detection.</p> <p>Our preliminary studies have shown a very good performance of the mass detection system consisting of 91 mammograms. The performance was initially 0.78-0.80 for the areas A_z under the ROC curves using the conventional neural network, and later being improved to A_z values of 0.84-0.89 when using the newly developed multiple circular path neural networks.</p>				
14. SUBJECT TERMS Breast Cancer, image registration, change detection, mass detection				15. NUMBER OF PAGES 67
				16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

3. TABLE OF CONTENTS

Front Cover	1
SF 298	2
Table of Contents	3
Introduction	4
Body	5-9
Appendices	10-

4. INTRODUCTION

Temporal change of mass lesions overtime is a key piece of information in computer-aided diagnosis of breast cancer and treatment monitoring, the **purpose** of the project is to develop an automatic change detection method to quantitatively extract the clinically important changes of suspicious lesions, upgrade the existing CAD system, and thus improve the clinical diagnosis of breast cancer. We will build a site model for each individual patient for monitoring the breast tissue changes and extend our current research on image registration and change detection to the early detection of breast cancer. **Specific aims** include: 1) registration and segmentation of deformable breast tissue structures across a series of mammograms; 2) construction of a site model of the mammogram for individual patients showing the locations of regions of interest and associated diagnostic information; 3) identification of clinically significant changes in both global and local mass areas within the breast; and 4) integration and evaluation of the developed techniques with existing CAD prototype. At conclusion of this project, we anticipate **achieving** the following: 1) establish a reliable technique of monitoring breast tissue changes associated with cancerous masses; 2) deliver a CAD prototype that can incorporate tissue change information from additional mammograms; 3) evaluate the merit of combining change detection and CAD for improved clinical diagnosis using multiple mammograms; and 4) acquire the experience necessary to explore multimodality imaging for unified detection, diagnosis and treatment assessment of breast cancer.

5. BODY-Annual Summary

The **long-term** goal of this career development project is to develop image guided diagnosis methodology through change detection in mammogram sequences for breast cancer detection. The research requires the knowledge of image analysis, image registration, change detection, and computational intelligence.

During the third year of this career development project, I continued a close research collaboration with Dr. Matthew Freedman (breast/lung imaging) at Georgetown University Lombardi Cancer Center, and further developed a strategic collaboration with Dr. Zsolt Szabo (nuclear medicine) at Johns Hopkins Medical Institutions (JHMI). Through them, I have learned more about breast cancer at both imaging and molecular levels. I have been appointed as an adjunct assistant professor of radiology by JHMI and served as a panel member for evaluating Molecular Imaging Center Awards (P20/P50) and Innovative Molecular Analysis Technologies (R21/R33) for the National Cancer Institute.

I continued serving as the director of *Computational Imaging and Informatics Laboratory*, which currently hosts one associate professor, one assistant professor, one visiting professor, and five full-time graduate students. Three of my graduate students are the recipients of CDMRP-BCRP Predoctoral Traineeships.

I have been promoted to the rank of associate professor after five-year faculty service at CUA.

As the research accomplishments during the third year, I have first identified the following major research tasks:

1. Develop a working algorithm for two-dimensional registration of image sequence data sets. It consists of three major components: (1) control points/objects extraction; (2) mixture of probabilistic principle axes registration (mPAR), and (3) deformable image warping through multilayer perceptron (MLP) neural network computations.
2. Develop new pattern recognition algorithm to perform computer-aided diagnosis for mass detection, where the lesion candidates are firstly identified by change detection using independent component analysis.

Follow this plan, major research accomplishments are described as follows:

5.1 Hybrid non-rigid image registration algorithm

We have developed a neural computation based non-rigid registration methodology using multiple rigid transforms, in a piece-wise fashion, to model the registration process between images in a sequence. The registration methodology is a hybrid approach that combines registration without exact point correspondence via multi-object principal axes, and registration with point correspondence via polynomial transform. Neural computation is used, for the first, to combine the derived individual principal axes solutions for each object in a committee machine formulation, and to obtain the polynomial transform based on extracted control points using a multilayer perceptron (MLP).

In our method, we present a neural computation based non-rigid registration using piece-wise rigid transformation. The novel feature is to align two point sets without needing to establish explicit point correspondences, where the derivation is realized by minimizing the relative entropy between the two point distributions resulting in a maximum likelihood estimate of the transformation matrix. A committee machine approach is used for recovering the transformational geometry of the non-rigid structures. That is

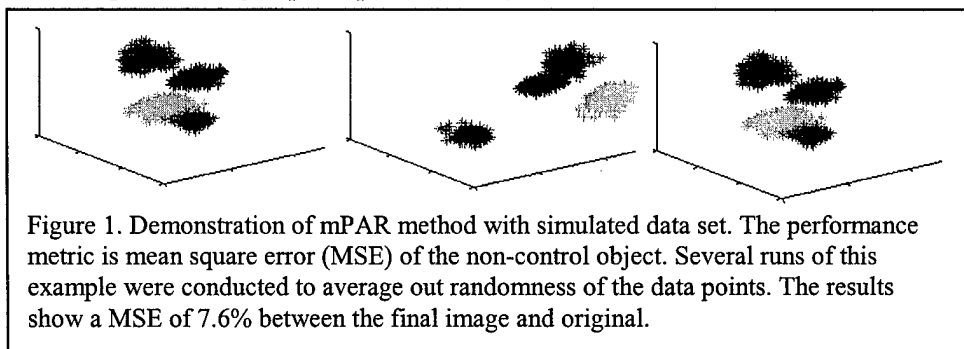


Figure 1. Demonstration of mPAR method with simulated data set. The performance metric is mean square error (MSE) of the non-control object. Several runs of this example were conducted to average out randomness of the data points. The results show a MSE of 7.6% between the final image and original.

That is

rather than using a single transformation matrix which gives rise to a large registration error, we attempt to interpolatively apply a mixture of transformations. By further generalizing PAR to a finite mixture registration (mPAR) scheme, with a soft partitioning of the data set, the mixture is fit using expectation-maximization (EM) algorithm. We then applied a probabilistic adaptive principal components extraction (PAPEX) algorithm, to estimate the transformational of the orthogonal set of eigenvalues and eigenvectors of the auto-covariance matrix. By applying a committee machine to a non-rigid registration, using FMR as the experts and PAPEX as a gating function, we can acquire the registration based on a mixture of piece-wise transformations of the data set. Then the correspondences control points are obtained. As a final step, the warped image is obtaining using the neural network based non-linear mapping, to obtain the polynomial transform based on extracted control points using MLP.

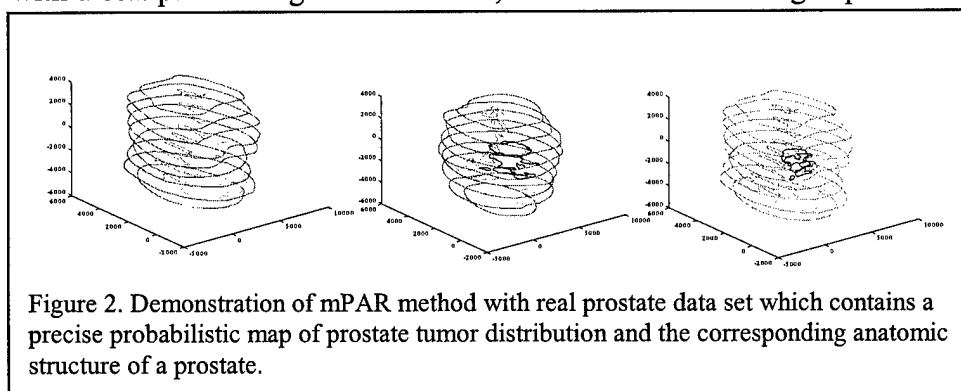


Figure 2. Demonstration of mPAR method with real prostate data set which contains a precise probabilistic map of prostate tumor distribution and the corresponding anatomic structure of a prostate.

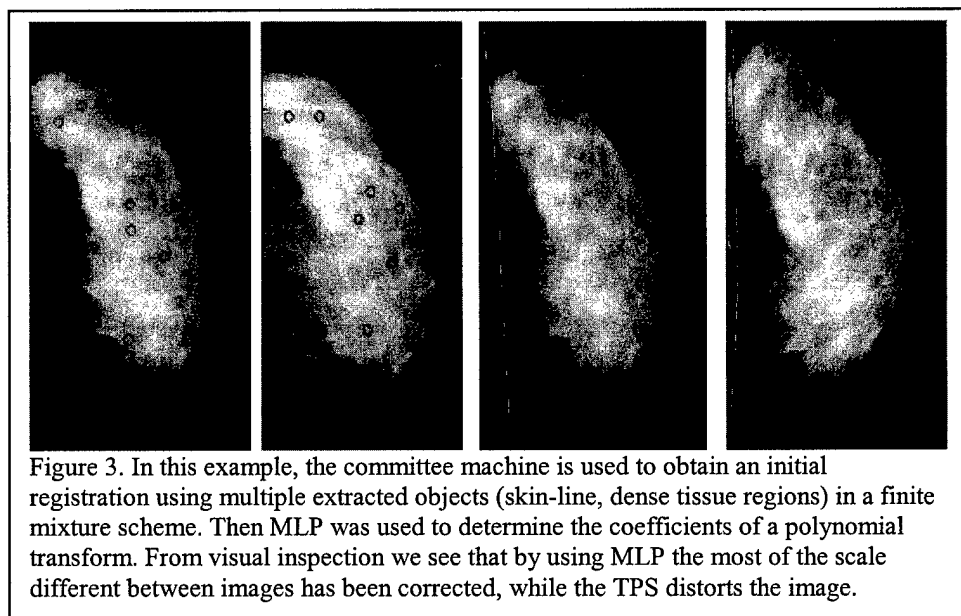


Figure 3. In this example, the committee machine is used to obtain an initial registration using multiple extracted objects (skin-line, dense tissue regions) in a finite mixture scheme. Then MLP was used to determine the coefficients of a polynomial transform. From visual inspection we see that by using MLP the most of the scale different between images has been corrected, while the TPS distorts the image.

Three examples are presented to demonstrate the techniques involved in the process. The first example uses four Gaussian clusters and focuses on the combination of the multiple transforms into a composite transform using finite mixture modeling techniques. The next examples present the complete process for prostate cancer registration and breast sequence analysis respectively. To verify performance, the results are compared to non-neural based implementations and other existing registration methods. **The major part of the work was published in IEEE Proceedings of Neural Network for Signal Processing, 2001 (attached).**

5.2 Feature-Based Computer-Aided Detection of Masses

In the clinical course of detecting masses, mammographers usually evaluate the surrounding background of a radiodense when breast cancer is suspected. In this study, we adapted this fundamental concept and computed features of the suspicious region in radial sections. These features were then arranged by circular convolution processes within a neural network, which led to an improvement in detecting mammographic masses.

In this study, randomly selected mammograms were processed by morphological enhancement techniques. Radiodense areas were isolated and delineated using a region growing algorithm with a valley blocking

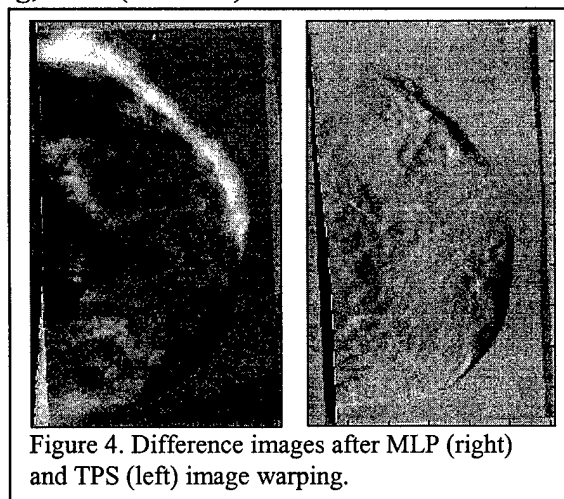


Figure 4. Difference images after MLP (right) and TPS (left) image warping.

technique. The boundary of each region of interest was then divided into 36 sectors using 36 equi-angle dividers radiated from the center of the area. Four features at each section were computed: (1) the radius, (2) the normal angle of the boundary, (3) the average gradient along the radial direction, and (4) the gray value difference (i.e., contrast) along the radial direction. Hence, 144 computed features (i.e., 4 features per sector for 36 sectors) were used as input values for the newly invented multiple circular path neural network (MCPNN). The neural network is constructed to emphasize on the correlation information associated with the feature interactions within the angle and between adjacent angles.

We have tested this approach on our research database consisting of 91 mammograms. The over-all performance in the detection of masses was 0.78-0.80 for the areas (Az) under the ROC curves using the conventional neural network. Later, the performance was improved to Az values of 0.84-0.89 using the multiple circular path neural networks. **The major part of the work was published in IEEE Transactions on Medical Imaging, 2001** (Two regular papers and one Technical Report are attached).

5.3 Key results and discussion

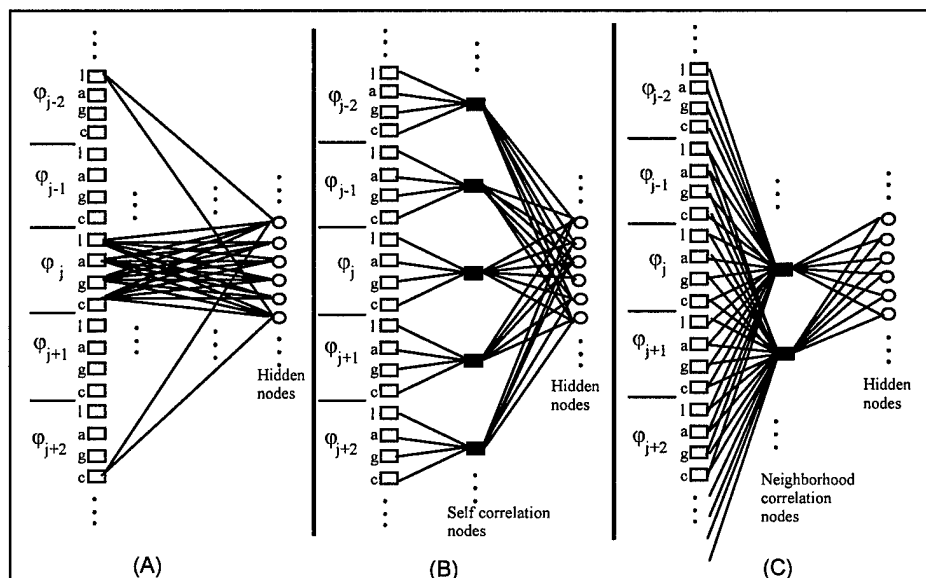


Figure 5. Three types of network paths connecting the input and the hidden layers: (1) Full connection. (2) A self correlation (SC) path; each node on the layer connects to a single set of the features (l,a,g,c) for the fan-in and fully connects to the hidden nodes for fan-out. (C) A neighborhood correlation (NC) path; each node on the layer connects to five adjacent sets of the features for the fan-in and fully connects to the hidden nodes for fan-out. Note that the fan-in nets emphasizing self correlation in (2) and neighborhood correlation in (3) represent convolution weights (i.e., the same type of sectors possess the same set of weighting factors).

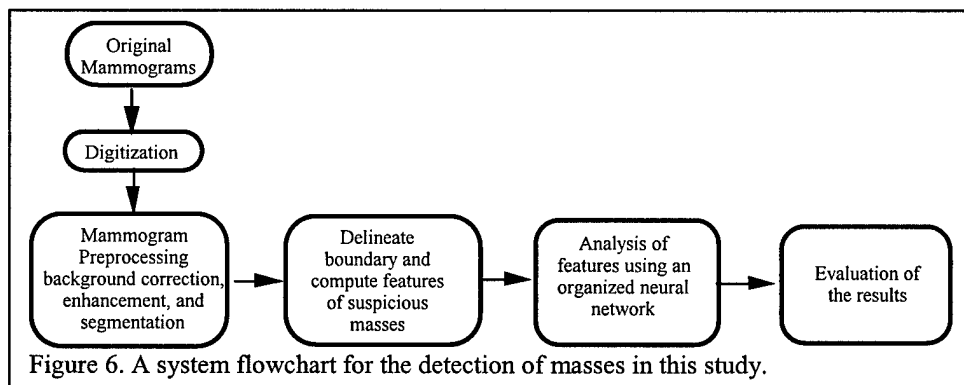


Figure 6. A system flowchart for the detection of masses in this study.

During the second year of the project, we worked on the theoretical concepts and methods of a neural computation based non-rigid registration algorithm. The approach uses a committee machine to recover the total transformational geometry of the non-rigid object using multiple rigid transforms combined together in a finite mixture registration scheme (mPAR). Finite mixture transform combination is a novel technique to combine multiple transforms that are contained in a single image. Different from local transforms, no other method combines multiple transforms together. In addition, finite mixture combinations, yields a lower MSE than the local transform and results in a smooth image while local transforms yield an image containing discontinues on transform boundaries. Furthermore, the registration obtained in the committee machine is fine tuned using a non-linear transform generated by a MLP network using extracted control points. **To the best of our knowledge, we may be the first group to develop neural networks based image registration techniques for nonlinear warping of image sequences.**

We applied our new algorithms to prostate and breast registration problems with reasonable results as shown in the previous figures. Some distortion can be seen in the final warped images because of the error in control point selection and correspondence.

Improvement in this portion should decrease distortion and yield a smoother looking image. Using neural networks in this problem has increased the generality of this approach by allowing the algorithm to adjust performance as imaging condition change.

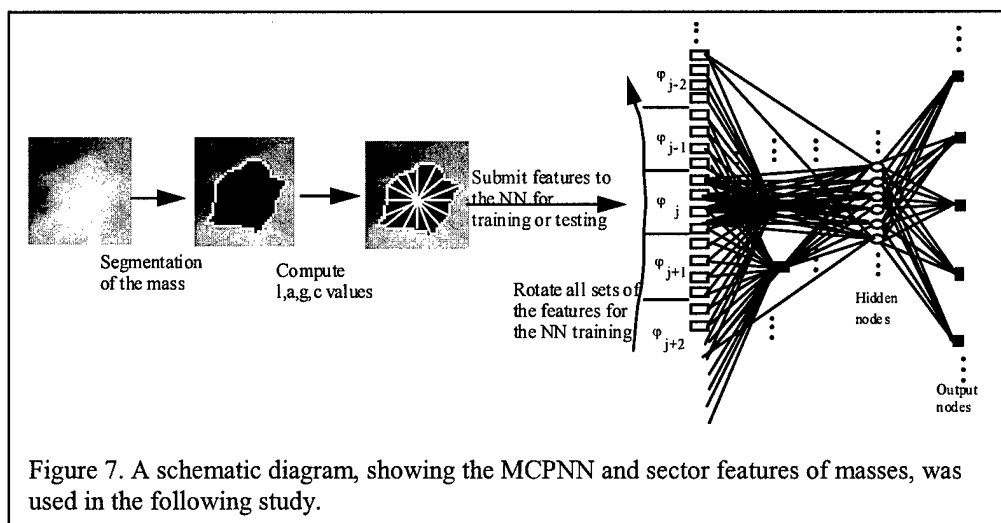


Figure 7. A schematic diagram, showing the MCPNN and sector features of masses, was used in the following study.

In neural network based mass detection, we randomly selected 54 computer-segmented areas where 30 patches were matched with the radiologist's identification and 24 were not. This database was used to train two neural network systems: (1) a conventional 3-layer BP neural network (with 125 nodes in the hidden layer) and (2) the proposed MCPNN training method using the same neural network learning algorithm. The structure of the MCPNN was described earlier. However, we used one fully connected path, four SC paths, four 20° NC paths, four 30° NC paths, three 40° NC paths, and two 50° NC paths in the first step network connection for the MCPNN. All paths in the neural network have their hidden layers. Only one hidden layer per path was used. Both neural network systems were trained by the error back propagation algorithm by feeding the features from the input layer and registering the corresponding target value at the output side. Once the training of the neural networks was complete, we then used the remaining 71 computer segmented areas for the testing. None of the images and their corresponding patients in the testing set could be found in the training set. The neural network output values were fed into the LABROC program for the performance evaluation. The results indicated that the areas (A_z) under the receiving operator characteristic (ROC) curves were 0.781 and 0.844 using the conventional BPNN and the MCPNN, respectively. The ROC curves of these two neural network training methods are shown in Figure 8 (left). We also invited another senior mammographer to conduct an ROC observer study. The mammographer was asked to rate each patch using a numerical scale ranging 0-10 for its likelihood of being a mass. These 71 numbers were also fed into the LABROC program. The mammographer's performance in A_z on this set of test cases was 0.909. The corresponding ROC curve is also shown in Figure 8 (left).

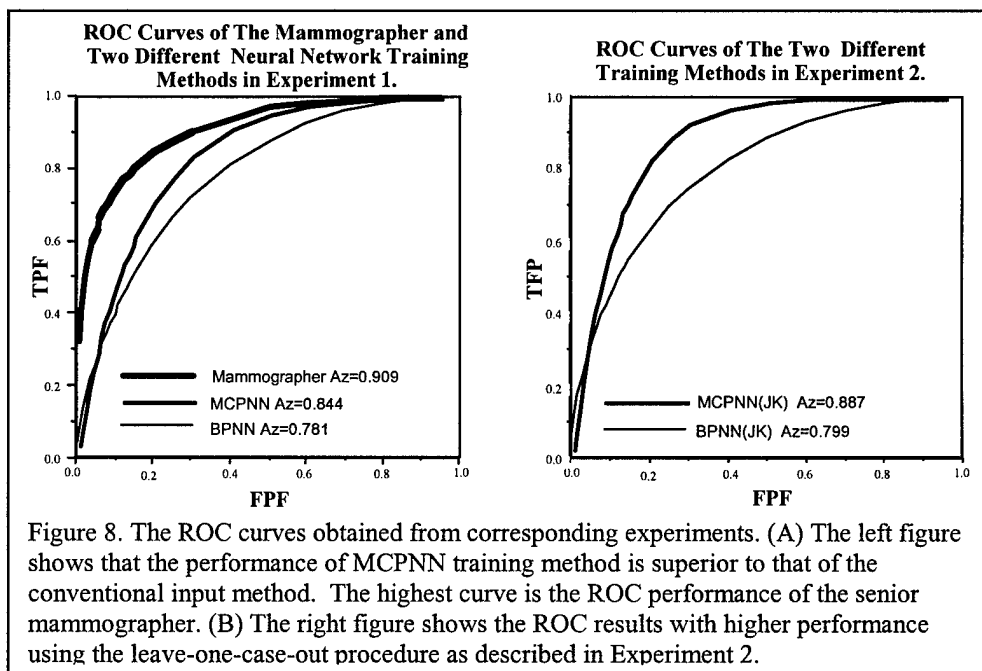


Figure 8. The ROC curves obtained from corresponding experiments. (A) The left figure shows that the performance of MCPNN training method is superior to that of the conventional input method. The highest curve is the ROC performance of the senior mammographer. (B) The right figure shows the ROC results with higher performance using the leave-one-case-out procedure as described in Experiment 2.

The neural network output values were fed into the LABROC program for the performance evaluation. The results indicated that the areas (A_z) under the receiving operator characteristic (ROC) curves were 0.781 and 0.844 using the conventional BPNN and the MCPNN, respectively. The ROC curves of these two neural network training methods are shown in Figure 8 (left). We also invited another senior mammographer to conduct an ROC observer study. The mammographer was asked to rate each patch using a numerical scale ranging 0-10 for its likelihood of being a mass. These 71 numbers were also fed into the LABROC program. The mammographer's performance in A_z on this set of test cases was 0.909. The corresponding ROC curve is also shown in Figure 8 (left).

We also conducted a leave-one-case-out experiment using the same database. In this experiment, we used those patches extracted from 90 mammograms for the training and used the patches (most of them are single) extracted from the remaining one mammogram as test objects. The procedure was repeated 91 times to allow every suspicious patch from each mammogram to be tested in the experiment. For each individual suspicious area, the computed features were identical to those used in Experiment 1. Again, both neural network systems were independently evaluated with the same procedure. The results indicated that the A_z values were 0.799 and 0.887 using the conventional back propagation neural network and the MCPNN, respectively. Figure 8 (right) shows the ROC curves of these two neural network systems using the leave-one-of-out procedure in the experiment.

Through this study, we found that the selected features are somewhat effective in the detection of masses. These features were "computationally translated" from the qualitative descriptors of BI-RAD. Another uniqueness of this study was on the test of our newly developed MCPNN training method. In Experiment 1, we found that the performances of both neural network systems were increased as compared to existing algorithms. This might be due to the increased number of cases (from 54 to 124) in the training set. In Experiment 2, the A_z value was improved by 0.043 using the MCPNN training method that was higher than A_z difference of 0.018 obtained by the conventional training method. The results implied that the MCPNN learned more effectively than the conventional MCPNN-BP when the number of training cases was increased.

It is known in the field of artificial intelligence that the key factors in pattern recognition are: (1) effective methods in the extraction of features and (2) analytic methods (e.g., back propagation neural network) for the extracted features. In this study, we showed that the training method designed to guide the analyzer is also an important factor to a success of a pattern recognition task. Though this finding is not new, the research of developing training methods for various pattern recognition tasks has not established in the field of medical imaging. In this work, we demonstrated that organized features with proper network connection and task-oriented guidance would assist the neural network in performing the task.

As far as the research in recognition of masses is concerned, we believe that main concept of using sectors is an effective approach. Note that any features arranged in the polar coordinate system can be trained by the MCPNN method. Since the MCPNN only coordinates the input data, the internal neural network learning algorithm can be changed to other learning algorithms. A technique using the rubber band straightening transformation, independently developed by Sahnier et al., for the detection of masses also employs a similar concept in extracting feature and/or texture in the polar coordinate. We believe that integration of effective feature and texture values computed at small sectors will be the research trend in mass detection. This paper focuses on neural network design and arrangement of features for effective pattern recognition of masses in medical imaging.

6. APPENDICES

6.1 Key Research Accomplishments

- We have proposed and developed mixture of principal axes registration (mPAR) algorithm to perform initial non-rigid image warping, without control point correspondence. A working computer code has been implemented. This work is original and new.
- We have proposed and developed multilayer perceptron (MLP) neural computation to perform final non-rigid image warping, with initially matched control points correspondence. A working computer code has been implemented. This work is original and new.
- Our numerical evaluation has shown that new image registration outperform existing popular methods such as Thin-Plate Spline (TSP) algorithm.
- We have proposed and developed multiple circular path neural network (MCPNN) classifier to perform computer-aided mass detection, against other existing methods. We have implemented related algorithms. This work is original and new.
- We have tested our detection system using receiver operating characteristics (ROC) analysis. Our preliminary experiment has shown that new system outperform existing popular methods such as conventional backpropagation (BP) neural networks.

6.2 Reportable Outcomes

- 1) H. Li, **Y. Wang**, K-J R. Liu, S-H B. Lo, and M. T. Freedman, "Computerized Radiographic Mass Detection-Part I: Lesion Site Selection by Morphological Enhancement and Contextual Segmentation," *IEEE Transactions on Medical Imaging*, vol. 20, no. 4, pp. 289-301, April 2001.
- 2) H. Li, **Y. Wang**, K-J R. Liu, S-H B. Lo, and M. T. Freedman, "Computerized Radiographic Mass Detection-Part II: Decision Support by Featured Database Visualization and Modular Neural Networks," *IEEE Transactions on Medical Imaging*, vol. 20, no. 4, pp. 302-313, April 2001.
- 3) **Y. Wang**, T. Adali, J. Xuan, and Z. Szabo, "Magnetic Resonance Image Analysis by Information Theoretic Criteria and Stochastic Site Models," *IEEE Transactions on Information Technology in Biomedicine*, vol. 5, no. 2, pp. 150-158, June 2001.
- 4) R. Srikanchana, K. Woods, J. Xuan, M. T. Freedman, C. Nguyen, and **Y. Wang**, "Non-Rigid Image Registration by Neural Computations," *Proceedings of IEEE Neural Networks for Signal Processing*, September 2001.
- 5) S-C B. Lo, H. Li, A. Hasegawa, **Y. Wang**, M. T. Freedman, and S. K. Mun, "A Multiple Circular Path Neural Network Architecture for Detection of Mammographic Masses," submitted to *IEEE Transactions on Medical Imaging*, 2001.

Computerized Radiographic Mass Detection—Part I: Lesion Site Selection by Morphological Enhancement and Contextual Segmentation

Huai Li, Yue Wang, K. J. Ray Liu*, Shih-Chung B. Lo, and Matthew T. Freedman

Abstract—This paper presents a statistical model supported approach for enhanced segmentation and extraction of suspicious mass areas from mammographic images. With an appropriate statistical description of various discriminate characteristics of both true and false candidates from the localized areas, an improved mass detection may be achieved in computer-assisted diagnosis (CAD). In this study, one type of morphological operation is derived to enhance disease patterns of suspected masses by cleaning up unrelated background clutters, and a model-based image segmentation is performed to localize the suspected mass areas using stochastic relaxation labeling scheme. We discuss the importance of model selection when a finite generalized Gaussian mixture is employed, and use the information theoretic criteria to determine the optimal model structure and parameters. Examples are presented to show the effectiveness of the proposed methods on mass lesion enhancement and segmentation when applied to mammographical images. Experimental results demonstrate that the proposed method achieves a very satisfactory performance as a preprocessing procedure for mass detection in CAD.

Index Terms—Finite mixture, image enhancement, image segmentation, information criterion, morphological filtering, relaxation labeling.

I. INTRODUCTION

IN RECENT years, several computer-assisted diagnosis (CAD) schemes for mass detection and classification have been developed [1]–[13]. Though it may be difficult to compare the relative performance of these methods, because the reported performance strongly depends on the degree of subtlety of masses in the selected database, accurate selection

of suspected masses is considered a critical and first step due to the variability of normal breast tissue and the lower contrast and ill-defined margins of masses [3], [6], and since no subtle masses should be missed before any further analysis.

A number of image processing techniques have been proposed to perform suspicious mass site selection. Kobatake *et al.* [1] proposed using a iris filter to detect tumors as suspicious regions with very weak contrast to their background. Sameti *et al.* [7] used fuzzy sets to partition the mammographic image data. Lau and Yin *et al.* independently proposed using bilateral-subtraction to determine possible mass locations [9], [13]. Some other investigators proposed using pixel-based feature segmentation of spiculated masses [4], [8]. Kegelmeyer has reported promising results for detecting spiculated tumors based on local edge characteristics and Laws texture features [8]. Karssemeijer *et al.* [4] proposed to identify stellate distortions by using the orientation map of line-like structures. Recently, Petrick *et al.* [6] proposed a two-stage adaptive density-weighted contrast enhancement filtering technique along with edge detection and morphological feature classification for automatic segmentation of potential masses. Kupinski and Giger [3] presented a radial gradient index-based algorithm and a probabilistic algorithm for seeded lesion segmentation.

Nevertheless, to our best knowledge, few work has been dedicated to improve the task of lesion site selection although it is indeed a very crucial step in CAD. Especially, few studies have used and justified model-based image processing techniques for unsupervised lesion site selection [11]. Zwiggelaar *et al.* developed a statistical model to describe and detect the abnormal pattern of linear structures of spiculated lesions [2]. In their work, the probability density function of the observation vectors for each class is assumed to be normal. We have experienced that the “normal” distribution for each class is not true. Li *et al.* proposed using a Markov random field model to extract suspicious masses for mass detection [11]. In their study, most of model parameters were chosen empirically, and the mammogram was segmented into three regions (background, fat, and parenchymal or tumors).

Stochastic model-based image segmentation is a technique for partitioning an image into distinctive meaningful regions based on the statistical properties of both gray level and context images. A good segmentation result would depend on suitable model selection for a specific image modality [16], [17] where model selection refers to the determination of both the number of image regions and the local statistical distributions of each region. Furthermore, a segmentation result would be improved

Manuscript received February 3, 1997; revised January 9, 2001. This work was supported in part by the Department of Defense under Grants DAMD17-98-1-8045 and DAMD17-96-1-6254 through a subcontract from University of Michigan, Ann Arbor, and by the National Science Foundation (NSF) under NYI Award MIP-9457397. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was M. Giger. Asterisk indicates corresponding author.

H. Li is with the Electrical Engineering Department and Institute for Systems Research, University of Maryland at College Park, College Park, MD 20742 USA. He is also with the Department of Radiology, Georgetown University Medical Center, Washington, DC 20007 USA.

Y. Wang is with the Department of Electrical Engineering and Computer Science, The Catholic University of America, Washington, DC 20064 USA. He is also with the Department of Radiology, Georgetown University Medical Center, Washington, DC 20007 USA.

*K. J. Ray Liu is with the Electrical Engineering Department and Institute for Systems Research University of Maryland at College Park, College Park, MD 20742 USA (e-mail: kjrlu@eng.umd.edu).

S.-C. B. Lo and M. T. Freedman are with the Department of Radiology, Georgetown University Medical Center, Washington, DC 20007 USA.

Publisher Item Identifier S 0278-0062(01)02831-2.

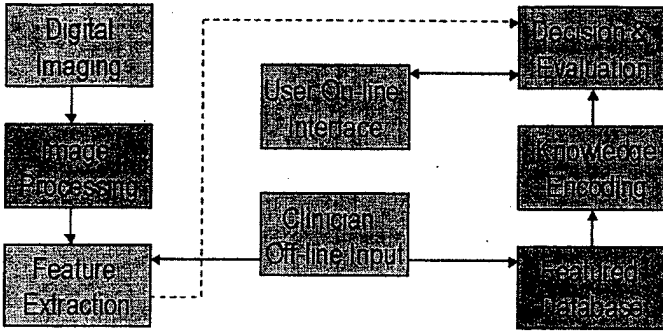


Fig. 1. Major components in CAD.

with preenhanced pattern of interest being segmented. The only assumption for suspected mass site selection is that suspected mass areas should be brighter than the surrounding breast tissues which is valid for most of the real cases. When some masses lie either within an inhomogeneous pattern of fibroglandular tissue or are partially or completely surrounded by fibroglandular tissue, enhancement of mass-related signals is important.

Fig. 1 shows a general block diagram of CAD systems. This paper focuses on "image processing" block, to just automatically pick up all possible lesion sites. We aim on two essential issues in the stochastic model-based image segmentation: enhancement and model selection. Based on the differential geometric characteristics of masses against the background tissues, we propose one type of morphological operation to enhance the mass patterns on mammograms. Then we employ a finite generalized Gaussian mixture (FGGM) distribution to model the histogram of the mammograms where the statistical properties of the pixel images are largely unknown and are to be incorporated. We incorporate the EM algorithm with two information theoretic criteria to determine the optimal number of image regions and the kernel shape in the FGGM model. Finally, we apply a contextual Bayesian relaxation labeling (CBRL) technique to perform the selection of suspected masses. The major differences of our work from the previous work [1]–[6], [8]–[13] are as follows.

- 1) We present a new algorithm of morphological filtering for image enhancement in which the combined operations are applied to the original gray tone image and the higher sensitive lesion site selection of the enhanced images are observed.
- 2) We justify and pilot test the FGGM distribution in modeling mammographic pixel images together with a model selection procedure based on the two information theoretic criteria. This allows an automatic identification of both the number (K) and kernel shape (α) of the distributions of tissue types.
- 3) We develop a new algorithm (CBRL) for segmenting mass areas where the comparable results are achieved as those using Markov random field model-based approaches while with much less computational complexity.

The presentation of this paper is organized as follows. In Section II, the proposed dual morphological operation enhancement technique is described in detail. The theory and algorithm on

FGGM modeling, model selection, and parameter estimation are presented in Section III. This is followed by a discussion on the selection of suspicious masses using the CBRL approach. Evaluation results are given and discussed in Section IV. Finally, the paper is concluded by Section V.

II. MORPHOLOGICAL ENHANCEMENT

One of the main difficulties in suspicious mass segmentation is that mammographic masses are often overlapped with dense breast tissues. Therefore, it is necessary to remove bright background caused by dense breast tissues while preserving the features and patterns related to the masses. For this purpose, background correction is an important step for mass segmentation. We propose a mass pattern-dependent background removal approach using morphological operations.

A. Morphological Filtering Theory

Morphological operations can be employed for many image processing purposes, including edge detection, region segmentation, and image enhancement. The beauty and simplicity of mathematical morphology approach come from the fact that a large class of filters can be represented as the combination of two simple operations: erosion and dilation. Let Z denote the set of integers and $f(i, j)$ denote a discrete image signal, where the domain set is given by $\{i, j\} \in N_1 \times N_2$, $N_1 \times N_2 \subset Z^2$ and the range set by $\{f\} \in N_3$, $N_3 \subset Z$. A structuring element B is a subset in Z^2 with a simple geometrical shape and size. Denote $B^s = \{-b : b \in B\}$ as the symmetric set of B and B_{t_1, t_2} as the translation of B by (t_1, t_2) , where $(t_1, t_2) \in Z^2$. The erosion $f \ominus B^s$ and dilation $f \oplus B^s$ can be expressed as [19]

$$(f \ominus B^s)(i, j) = \min_{t_1, t_2 \in B_{i,j}} (f(t_1, t_2)) \quad (1)$$

$$(f \oplus B^s)(i, j) = \max_{t_1, t_2 \in B_{i,j}} (f(t_1, t_2)). \quad (2)$$

On the other hand, opening $f \circ B$ and closing $f \bullet B$ are defined as [19]

$$(f \circ B)(i, j) = ((f \ominus B^s) \oplus B)(i, j) \quad (3)$$

$$(f \bullet B)(i, j) = ((f \oplus B^s) \ominus B)(i, j). \quad (4)$$

A gray value image can be viewed as a two-dimensional surface in a three-dimensional space. Given an image, the opening operation removes the objects, which have size smaller than the structuring element, with positive intensity. Thus, with the specified structuring element, one can extract different image contexts by taking the difference between the original and opening processed image, which is known as "tophat" operation [19].

B. Morphological Enhancement Algorithms

Based on the properties of morphological filters, we designed one type of mass pattern-dependent enhancement approaches. The algorithm is implemented by dual morphological tophat operations following by a subtraction which is described as follows.

Step 1) The textures without the pattern information of interest are extracted by a tophat operation

$$r_1(i, j) = \max(0, [f(i, j) - (f \circ B_1)(i, j)]) \quad (5)$$

where $f(i, j)$ is the original image, and $r_1(i, j)$ is the residue image between the original image and the opening of the original image by a specified structuring element B_1 . The size of B_1 should be chosen smaller than the size of masses.

Step 2) Let $r_2(i, j)$ be the mass pattern enhanced image by background correction, i.e., by the second tophat operation on $f(i, j)$

$$r_2(i, j) = \max(0, [f(i, j) - (f \circ B_2)(i, j)]) \quad (6)$$

where B_2 is a specified structuring element which has a larger size than masses.

Step 3) The enhanced image $f_1(i, j)$ can be derived as

$$f_1(i, j) = \max(0, [r_2(i, j) - r_1(i, j)]) \quad (7)$$

This operation is called "dual morphological operation." It can remove the background noise and the structure noise inside the suspected mass patterns. Fig. 2 shows the mass patch and the enhanced results of each step using the dual morphological operation. As we can see from Fig. 2, both background correction [Fig. 2(c)] and dual morphological operation [Fig. 2(d)] enhanced the mass pattern, but dual morphological operation removed more structural noise inside the mass region which in turn would improve the mass segmentation results.

III. MODEL-BASED SEGMENTATION

A. Statistical Modeling

Given a digital image consisting of $N_1 \times N_2$ pixels, assume this image contains K regions. By randomly reordering all pixels in the underlying probability space, one can treat pixel labels as random variables and introduce a prior probability measure π_k . Then the FGGM probability density function (pdf) of gray level of each pixel is given by [17]

$$p(x_i) = \sum_{k=1}^K \pi_k p_k(x_i), \quad i = 1, \dots, N_1 N_2, \quad (8)$$

$$x_i = 0, 1, \dots, L - 1$$

where x_i is the gray level of pixel i , and L is the number of gray levels. $p_k(x_i)$ s are conditional region pdfs with the weighting factor π_k , satisfying $\pi_k > 0$, and $\sum_{k=1}^K \pi_k = 1$. The generalized Gaussian pdf given region k is defined by

$$p_k(x_i) = \frac{\alpha \beta_k}{2\Gamma(1/\alpha)} \exp[-|\beta_k(x_i - \mu_k)|^\alpha], \quad \alpha > 0, \quad (9)$$

$$\beta_k = \frac{1}{\sigma_k} \left[\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)} \right]^{1/2}$$

where μ_k is the mean, $\Gamma(\cdot)$ is the Gamma function. β_k is a parameter related to the variance σ_k . It can be shown that when

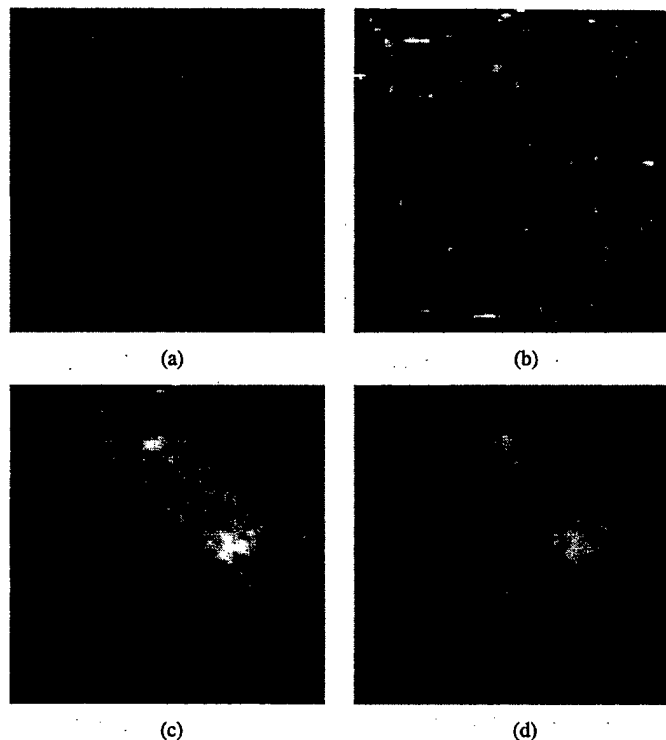


Fig. 2. Original and enhancement result of the mass patch using dual-morphological operation. (a) Original image block $f(i, j)$. (b) Textures $r_1(i, j)$. (c) Background correction result $r_2(i, j)$. (d) Enhanced result $f_1(i, j)$.

$\alpha = 2.0$, one has the Gaussian pdf; when $\alpha = 1.0$, one has the Laplacian pdf. When $\alpha \gg 1$, the distribution tends to a uniform pdf; when $\alpha < 1$, the pdf becomes sharp. Therefore, the generalized Gaussian model is a suitable model to fit the histogram distribution of those images whose statistical properties are unknown since the kernel shape can be controlled by selecting different α values.

The whole image can be well approximated by an independent and identically distributed random field \mathbf{X} . The corresponding joint pdf is

$$P(\mathbf{x}) = \prod_{i=1}^{N_1 N_2} \sum_{k=1}^K \pi_k p_k(x_i) \quad (10)$$

where $\mathbf{x} = [x_1, x_2, \dots, x_{N_1 N_2}]$, and $\mathbf{x} \in \mathbf{X}$. $p_k(x_i)$ is given in (9). Based on the joint probability measure of pixel images, the likelihood function under FGGM modeling can be expressed as $\mathcal{L}(\mathbf{r}) = \prod_{i=1}^{N_1 N_2} p_{\mathbf{r}}(x_i)$ where $\mathbf{r} : \{K, \alpha, \pi_k, \mu_k, \sigma_k, k = 1, \dots, K\}$ denotes the model parameter set.

B. Model Identification

With an appropriate system likelihood function, the objective of model identification is to estimate the model parameters by maximizing the likelihood function, or equivalently minimizing the relative entropy between the image histogram $p_{\mathbf{x}}(u)$ and the estimated pdf $p_{\mathbf{r}}(u)$, where u is the gray level. Based on the FGGM model, the EM algorithm is applied to estimate the model parameters. The EM algorithm is an iterative technique for maximum-likelihood (ML) estimation [20]. Recently, it has been used in many medical imaging applications [15]. Instead

of evaluating directly the value of ML, we use the global relative entropy (GRE) between the histogram and the estimated FGGM distribution to measure the performance of parameter estimation, given by

$$\text{GRE}(p_x||p_r) = \sum_u p_x(u) \log \frac{p_x(u)}{p_r(u)}. \quad (11)$$

Motivated by the same spirit of conventional EM algorithm for finite normal mixtures (FNMs), we formulated the EM algorithm to estimate the parameter values of the FGGM. The algorithm is summarized as follows.

EM Algorithm:

- 1) For $\alpha = \alpha_{\min}, \dots, \alpha_{\max}$
 - $m = 0$, given initialized $r^{(0)}$
 - E-step: for $i = 1, \dots, N_1 N_2$, $k = 1, \dots, K$, compute the probabilistic membership

$$z_{ik}^{(m)} = \frac{\pi_k^{(m)} p_k(x_i)}{\sum_{k=1}^K \pi_k^{(m)} p_k(x_i)}. \quad (12)$$

- M-step: for $k = 1, \dots, K$, compute the updated parameter estimates

$$\begin{cases} \pi_k^{(m+1)} = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1 N_2} z_{ik}^{(m)} \\ \mu_k^{(m+1)} = \frac{1}{N_1 N_2 \pi_k^{(m+1)}} \sum_{i=1}^{N_1 N_2} z_{ik}^{(m)} x_i \\ \sigma_k^{2(m+1)} = \frac{1}{N_1 N_2 \pi_k^{(m+1)}} \sum_{i=1}^{N_1 N_2} z_{ik}^{(m)} (x_i - \mu_k^{(m+1)})^2 \end{cases} \quad (13)$$

- When $|\text{GRE}^{(m)}(p_x||p_r) - \text{GRE}^{(m+1)}(p_x||p_r)| \leq \epsilon$ is satisfied, go to Step 2 Otherwise, $m = m + 1$ and go to E-Step.

- 2) Compute GRE, and go to Step 1.
- 3) Choose the optimal \hat{r} which corresponds to the minimum GRE.

As we mentioned in Section I, the two important parameters in model selection are K and α . Determination of the region parameter K directly affects the quality of the resulting model parameter estimation and in turn, affects the result of segmentation. In this paper we propose an approach to determine the value of K based on two popular information theoretic criteria introduced by Akaike [23] and by Rissanen [24]. Akaike proposed to select the model that gives the minimum Akaike information criterion (AIC), defined by

$$\text{AIC}(K) = -2 \log(\mathcal{L}(\hat{r}_{ML})) + 2K' \quad (14)$$

where \hat{r}_{ML} is the ML estimate of the model parameter set r , and K' is the number of free adjustable parameters in the model [15], [23]. AIC criterion will select the correct number of the image regions K_0 when

$$K_0 = \arg \left\{ \min_{1 \leq K \leq K_{\max}} \text{AIC}(K) \right\}. \quad (15)$$

Rissanen addressed the problem from a quite different point of view. Rissanen reformulated the problem explicitly as an information coding problem in which the best model fitness is measured such that it assigns high probabilities to the observed data while at the same time the model itself is not too complex to describe [24]. The model is selected by minimizing the total description length defined by minimum description length (MDL)

$$\text{MDL}(K) = -\log(\mathcal{L}(\hat{r}_{ML})) + 0.5K' \log(N_1 N_2). \quad (16)$$

Similarly, the correct number of the distinctive image regions K_0 will be estimated when

$$K_0 = \arg \left\{ \min_{1 \leq K \leq K_{\max}} \text{MDL}(K) \right\}. \quad (17)$$

C. Bayesian Relaxation Labeling

Once the FGGM model is given, a segmentation problem is the assignment of labels to each pixel in the image. A straightforward way is to label pixels into different regions by maximizing the individual likelihood function $p_k(x)$. This approach is called ML classifier, which is equivalent to a multiple thresholding method. Usually, this method may not achieve a good performance since there is lack of local neighborhood information to be included to make a good decision. CBRL algorithm [25] is one of the approaches, which can incorporate the local neighborhood information into labeling procedure and thus improve the segmentation performance. In this study, we developed the CBRL algorithm to perform/refine pixel labeling based on the localized FGGM model, which is defined as follows.

Let ∂i be the neighborhood of pixel i with an $m \times m$ template centered at pixel i . An indicator function is used to represent the local neighborhood constraints $R_{ij}(l_i, l_j) = I(l_i, l_j)$, where l_i and l_j are labels of pixels i and j , respectively. Note that pairs of labels are now either compatible or incompatible. Similar to reference [25], one can compute the frequency of neighbors of pixel i which has the same label values k as at pixel i

$$\pi_k^{(i)} = p(l_i = k | l_{\partial i}) = \frac{1}{m^2 - 1} \sum_{j \in \partial i, j \neq i} I(k, l_j) \quad (18)$$

where $l_{\partial i}$ denotes the labels of the neighbors of pixel i . Since $\pi_k^{(i)}$ is a conditional probability of a region, the localized FGGM pdf of gray level x_i at pixel i is given by

$$p(x_i | l_{\partial i}) = \sum_{k=1}^K \pi_k^{(i)} p_k(x_i) \quad (19)$$

where $p_k(x_i)$ is given in (9). Assuming gray values of the image are conditional independent, the joint pdf of x , given the context labels l , is

$$P(x|l) = \prod_{i=1}^{N_1 N_2} \sum_{k=1}^K \pi_k^{(i)} p_k(x_i) \quad (20)$$

where $l = (l_i : i = 1, \dots, N_1 N_2)$.

It is known that CBRL algorithm can obtain a consistent labeling solution based on the localized FGGM model (19). Since

TABLE I
DISTRIBUTION OF THE EFFECTIVE SIZE OF THE 186 MASSES USED IN THIS STUDY. THE EFFECTIVE SIZE IS DEFINED AS THE SQUARE ROOT OF THE PRODUCT OF THE MAXIMUM AND MINIMUM DIAMETERS OF THE MASS

	0–5mm	6–10mm	11–15mm	16–20mm	21–25mm	26–30mm
#	3	55	78	29	17	4

l represents the labeled image, it is consistent if $S_i(l_i) \geq S_i(k)$, for all $k = 1, \dots, K$ and for $i = 1, \dots, N_1 N_2$ [25], where

$$S_i(k) = \pi_k^{(i)} p_k(x_i). \quad (21)$$

Now we can define

$$A(l) = \sum_{i=1}^{N_1 N_2} \left(\sum_k I(l_i, k) S_i(k) \right) \quad (22)$$

as the average measure of local consistency, and

$$LC_i = \sum_k I(l_i, k) S_i(k), \quad i = 1, \dots, N_1 N_2 \quad (23)$$

represents the local consistency based on l . The goal is to find a consistent labeling l which can maximize (22). In the real application, each local consistency measure LC_i can be maximized independently. In [25], it has been shown that when $R_{ij}(l_i, l_j) = R_{ji}(l_j, l_i)$, if $A(l)$ attains a local maximum at l , then l is a consistent labeling.

Based on the localized FGGM model, $l_i^{(0)}$ can be initialized by ML classifier

$$l_i^{(0)} = \arg \left\{ \max_k p_k(x_i) \right\}, \quad k = 1, \dots, K. \quad (24)$$

Then, the order of pixels is randomly permuted and each label l_i is updated to maximize LC_i , i.e., classify pixel i into k th region if

$$l_i = \arg \left\{ \max_k \pi_k^{(i)} p_k(x_i) \right\}, \quad k = 1, \dots, K \quad (25)$$

where $p_k(x_i)$ is given in (9), $\pi_k^{(i)}$ is given in (18). By considering (24) and (25), we developed a modified CBRL algorithm as follows.

CBRL Algorithm:

- 1) Given $l^{(0)}$, $m = 0$
- 2) Update pixel labels
 - Randomly visit each pixel for $i = 1, \dots, N_1 N_2$
 - Update its label l_i according to

$$l_i^{(m)} = \arg \left\{ \max_k \pi_k^{(i)(m)} p_k(x_i) \right\}.$$

- 3) When

$$\frac{\sum (l^{(m+1)} \oplus l^{(m)})}{N_1 N_2} \leq 1\%,$$

stop; otherwise, $m = m + 1$, and repeat Step 2.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we present the results of using the morphological filtering and model-based segmentation approach we have introduced for enhancement and segmentation of suspi-

cious masses in mammographic images. In addition to the qualitative assessment by the radiologists, we introduce several objective measures to assess the performance of the algorithms we have proposed for enhancement and segmentation.

A testing data set of 200 mammograms and two simulated tone images were used to test and evaluate the performance of the algorithms in this study. The mammograms were selected from the Mammographic Image Analysis Society (MIAS) database and the Brook Army Medical Center (BAMC) database created by the Department of Radiology at Georgetown University Medical Center. Of the 200 mammograms, 50 mammograms are normal, and each of the 150 abnormal mammograms contains at least one mass case of varying size, subtlety, and location. The areas of suspicious masses were identified by an expert radiologist based on visual criteria and biopsy proven results. The total data set includes 113 benign and 73 malignant masses. The distribution of the masses in terms of size is shown in Table I. The BAMC films were digitized with a laser film digitizer (Lumiscan 150) at a pixel size of $100 \mu\text{m} \times 100 \mu\text{m}$ and 4096 gray levels (12 bits). Before the method was applied the digital mammograms were smoothed by averaging 4×4 pixels into one pixel. According to radiologists, the size of small masses is 3–15 mm in effective diameter. A 3-mm object in an original mammogram occupies 30 pixels in a digitized image with a $100\text{-}\mu\text{m}$ resolution. After reducing the image size by four times, the object will occupy the range of about seven to eight pixels. The object with the size of seven pixels is expected to be detectable by any computer algorithm. Therefore, the shrinking step is applicable for mass cases and can save computation time.

Experimental Evaluation of Morphological Enhancement: In order to justify the suitability of morphological structural elements, the geometric properties of the contexts and textures in mammograms were studied. The basic idea is to keep all mass-like objects within certain size range and remove all others by using the proposed morphological filters with specific structural elements. At the resolution of $400 \mu\text{m}$, a disk with a diameter of seven pixels was chosen as the morphological structuring elements B_1 to extract textures in mammograms. Since the smallest masses have seven pixels in diameter with the resolution of $400 \mu\text{m}$, this procedure would not destroy mass information. For the purpose of background correction, a disk with a diameter of 75 pixels was used as the morphological structuring element B_2 . An object with a diameter of 75 pixels corresponds to 30 mm in the original mammogram. This indicates that all masses with sizes up to 30 mm can be enhanced by background correction. Masses larger than 30 mm are rare cases in the clinical setting. In the last stage of our approach, we applied morphological opening and closing filtering using a disk with a diameter of five to eliminate small objects which also contribute to texture noise.

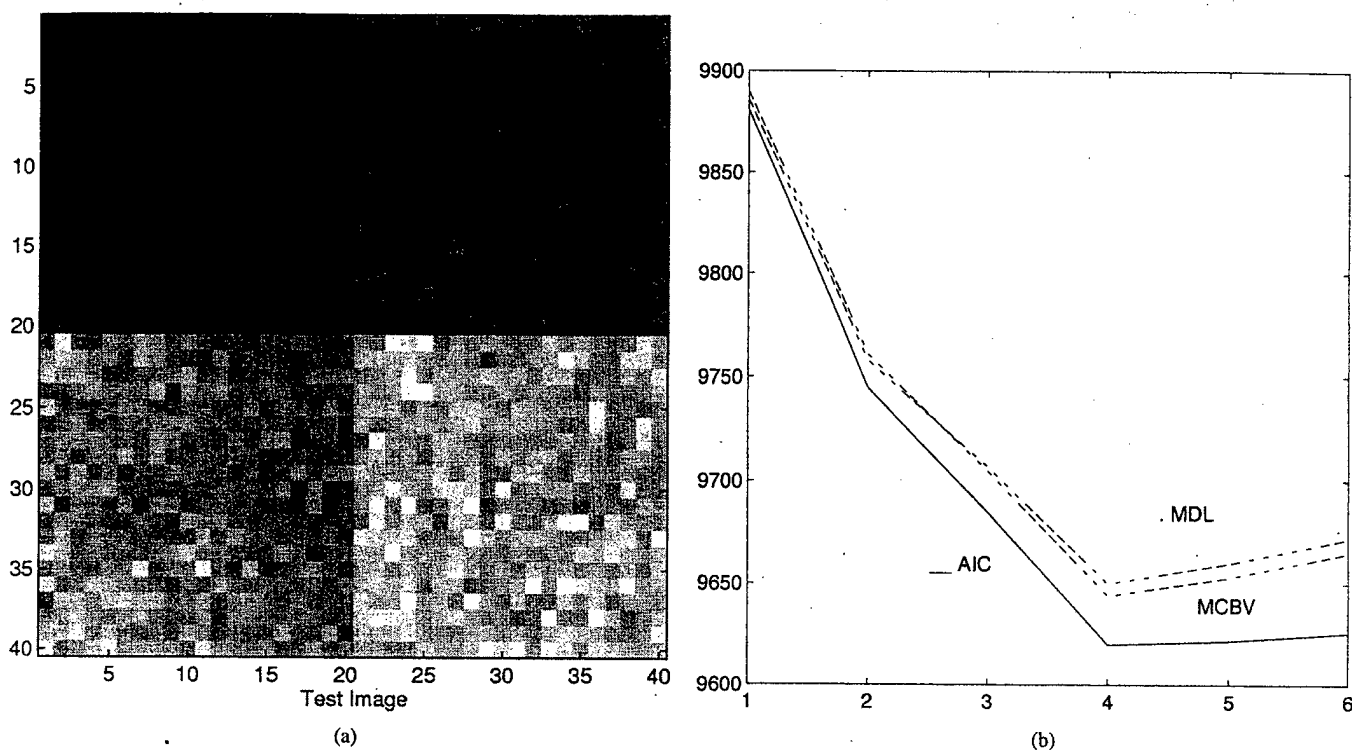


Fig. 3. (a) Original simulated test image for model selection ($k_0 = 4$, $\text{SNR} = 10$ dB) and (b) the AIC/MDL curves in model selection ($\sigma = 30$).

All testing mammograms were processed using the proposed enhancement approach with the suggested structuring element B_1 and B_2 . Fig. 5 shows processed mammogram examples using the morphological enhancement. Compared the enhanced results [Fig. 5(b) and (d)] with the original mammograms [Fig. 5(a) and (c)], the proposed method not only enhanced all suspected mass patterns and reduced the texture noise, but also removed the background noise. In summary, the proposed morphological enhancement approach can enhance mass patterns and remove texture structure noises. For dense mammograms, such as the second example in Fig. 5(c) and (d), the mass is obscured by dense fibroglandular tissues, our experience shows applying the dual morphological operation to remove the fibroglandular tissue background is useful. In addition to the visual evaluation by the radiologist, we performed the segmentation to assess the effectiveness of the morphological filtering, based on the enhanced mammograms and the original mammograms.

Simulated Evaluation of Segmentation Algorithms: The performance of model selection using two frequently used methods, i.e., the AIC and MDL [22], were first tested and compared in the simulation study. The computer-generated data was made up of four overlapping normal components. Each component represents one local region. The value for each component were set to a constant value, the noise of normal distribution was then added to this simulation digital phantom. Three noise levels with different variance were set to keep the same signal-to-noise ratio (SNR), where SNR is defined by

$$\text{SNR} = 10 \log_{10} \frac{(\Delta\mu)^2}{\sigma^2} \quad (26)$$

where $\Delta\mu$ is the mean difference between regions, and σ^2 is the noise power. The original data for the simulation study are

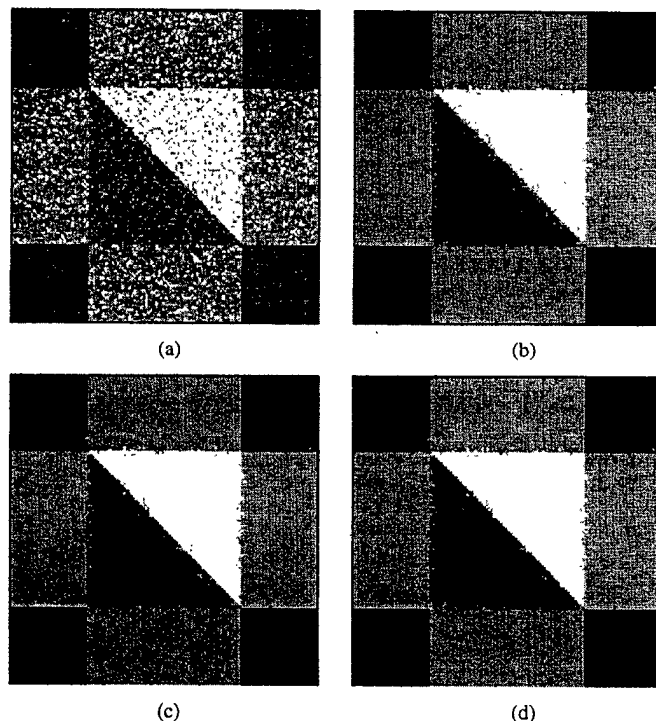


Fig. 4. Image segmentation by CBRL on simulated image (with initialization by ML classification). (a) ML initialization. (b) First iteration in CBRL. (c) Second iteration in CBRL. (d) Third iteration in CBRL.

TABLE II
COMPARISON OF CBRL, ICM, AND MICM ALGORITHM: SIMULATED DATA

Item	CBRL Result	ICM Result	MICM Result
Classification Error	0.7935%	0.7508%	0.3113%

given in Fig. 3(a). The AIC and MDL curves, as functions of the number of local clusters K , are plotted in Fig. 3(b). According

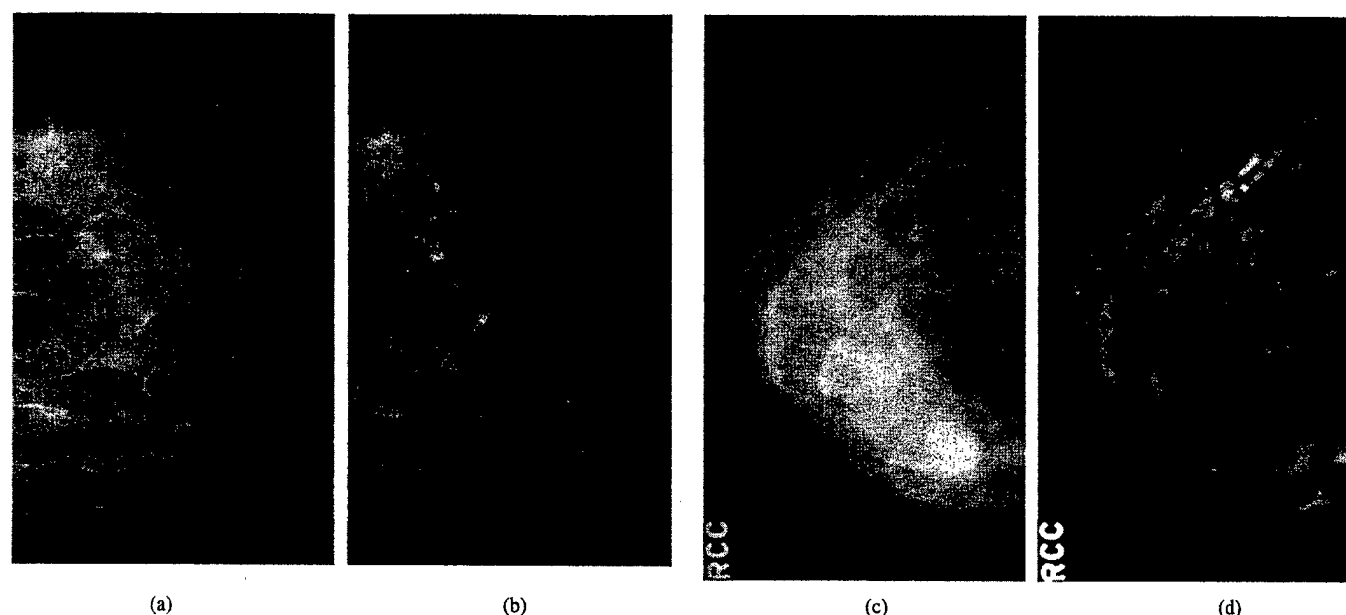


Fig. 5. Examples of mass enhancement. (a) Original mammogram. (b) Enhanced mammogram. (c) and (d) Another original mammogram and its enhanced result.

to the information theoretic criteria, the minima of these curves indicate the correct number of the local regions. From this experimental figure, it is clear that the number of local regions suggested by these criteria are all correct.

For the validation of image segmentation using CBRL, we apply the algorithm first to a simulated image. We use ML classifier to initialize image segmentation, i.e., to initialize the quantified image by selecting the pixel label with largest likelihood at each node. The classification error after initialization is uniformly distributed over the spatial domain as shown in Fig. 4(a). Our experience suggested this to be a very suitable starting point for contextual relaxation labeling [21]. The CBRL is then performed to fine tune the image segmentation. It should be emphasized that the ground truth is known in this simulated experiment, the percentage of total classification error is used as the criterion for evaluating the performance of segmentation technique. In Fig. 4(a)–(d), the initial segmentation by the ML classification and the stepwise results of three iterations in the CBRL are presented. In this experiment, algorithm initialization results in an average classification error of 30%. It can be clearly seen that a dramatic improvement is obtained after several iterations of the CBRL by using local constraints determined by the context information. In addition, the convergence is fast as one can see, after the first iteration most of the misclassification are removed. We have also implemented two other independent and popular algorithms, namely, the iterated conditional mode (ICM) and the modified iterated conditional mode (MICM) algorithms, so as to assess the comparative performance of the segmentation results among different approaches [21], [22]. The only assumption being made by these three methods is the Markovian property of the context images which can be well justified by the underlying cell oncology and pathology. We have applied these three algorithms to the same testing image and the corresponding classification errors are presented in Table II. The final percentage of classification errors for Fig. 4(d) is 0.7935%. From this experimental comparison, it can be concluded that three algorithms achieved com-

TABLE III
COMPUTED AICs FOR THE FGGM MODEL WITH DIFFERENT α

K	$\alpha = 1.0$	$\alpha = 2.0$	$\alpha = 3.0$	$\alpha = 4.0$
2	651250	650570	650600	650630
3	646220	644770	645280	646200
4	645760	644720	645260	646060
5	645760	644700	645120	646040
6	645740	644670	645110	645990
7	645640	644600	645090	645900
8	645550(min)	644570(min)	645030(min)	645850(min)
9	645580	644590	645080	645880
10	645620	644600	645100	645910

TABLE IV
COMPUTED MDLs FOR THE FGGM MODEL WITH DIFFERENT α

K	$\alpha = 1.0$	$\alpha = 2.0$	$\alpha = 3.0$	$\alpha = 4.0$
2	651270	650590	650630	650660
3	646260	644810	645360	646350
4	645860	644770	645280	646150
5	645850	644770	645280	646100
6	645790	644750	645150	646090
7	645720	644700	645120	645930
8	645680(min)	644690(min)	645100(min)	645900(min)
9	645710	644710	645140	645930
10	645790	644750	645180	645960

parable segmentation accuracy and the result produced by the MICM algorithm is most superior, though in terms of computational complexity the CBRL algorithm is the least. It should be noticed that since in MICM algorithm an inhomogeneous configuration of the Markov random field is used, its superior performance is reasonable.

On Model-Based Segmentation—Real Case Study: In the real case study, we used two information criteria (AIC and MDL) to determine K . Tables III and IV shows the AIC and MDL values with different K and α of the FGGM model based on one original mammogram. As it can be seen from Tables III and IV, although with different α , all AIC and MDL values achieve the minimum when $K = 8$. It indicates that AIC and

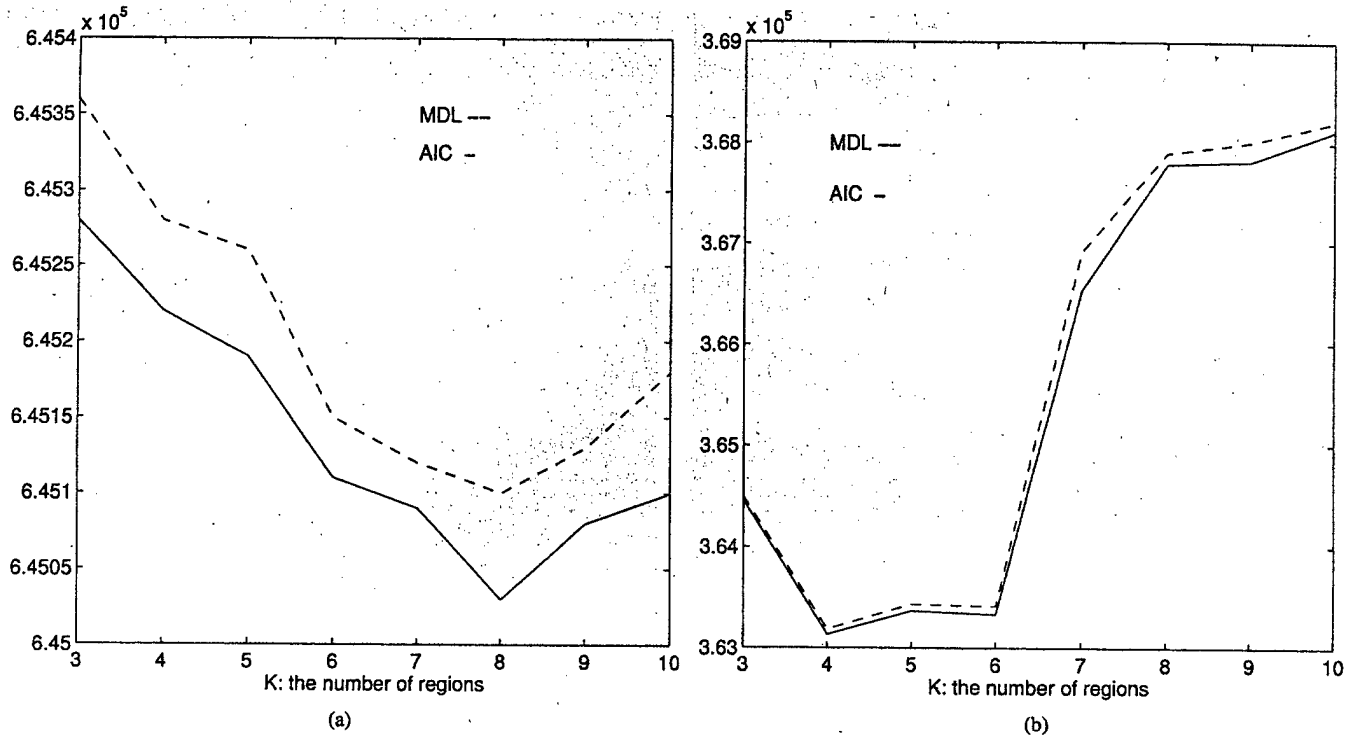


Fig. 6. AIC and MDL curves with different number of region K . (a) Result based on the original mammogram, the optimal $K = 8$. (b) Result based on the enhanced mammogram, the optimal $K = 4$.

MDL are relatively insensitive to the change of α . With this observation, we can decouple the relation between K and α and choose the appropriate value of one while fixing the value of another. Fig. 6(a) and (b) are two examples of AIC and MDL curves with different K and fixed $\alpha = 3.0$. Fig. 6(a) is based on the original mammogram and Fig. 6(b) is based on the enhanced mammogram. As we can see in Fig. 6(a), both criteria achieved the minimum when $K = 8$. It should be noticed that though no ground truth is available in this case, our extensive numerical experiments have shown a very consistent performance of the model selection procedure and all the conclusions were strongly supported by the previous independent work reported by [14]. Fig. 6(b) indicates that $K = 4$ is the appropriate choice for the mammogram enhanced by dual morphological operation. This is believed to be reasonable since the number of regions decrease after background correction.

We fixed $K = 8$, and changed the value of α for estimating the FGGM model parameters using the proposed EM algorithm with the original mammogram. The GRE value between the histogram and the estimated FGGM distribution was used as a measure of the estimation bias. We found that GRE achieved a minimum distance when the FGGM parameter $\alpha = 3.0$ as shown in Fig. 7. The similar result was shown when we applied the EM algorithm to the enhanced mammogram with $K = 4$. This indicated that the FGGM model might be better than the FNM model ($\alpha = 2.0$) in modeling mammographic images when the true statistical properties of mammograms are generally unknown, though the FNM has been most often chosen in many previous work [15].

After the determination of all model parameters, every pixel of the image was labeled to a different region (from 1 to K) based on the CBRL algorithm. We then selected the brightest re-

TABLE V
COMPARISON OF SEGMENTATION ERROR RESULTING FROM NONCONTEXTUAL AND CONTEXTUAL METHODS

Method	Soft Classification	Bayesian Classification	CBRL
GRE Value	0.0067	0.4406	0.1578

gion, which corresponding to label K , plus a criterion of closed isolated area, as the candidate region of suspicious masses. According to the visual inspections by the radiologists, when we use $K - 1$ instead of K , the results are over-segmented. For the case of using $K + 1$, the results are under-segmented. In order to quantify the performance differences between the different segmentation methods, several groups have suggested that the segmentation results may be compared against radiologists' outlines of the lesions [3]. Though the proposed comparison measures are quantitative, the performance measures are still qualitative, since the reference base (e.g., gold standard by the radiologists) is qualitative, subjective, and imperfect. Therefore, in this model-supported approach, in addition to the visual inspections by the radiologists, we have also introduced an objective measure, the GRE between the histogram of the pixel images $p_x(u)$ and the FGGM of the segmented image $p_{x,1}(u)$ to assess the performance of the segmentation, defined by

$$\text{GRE}(p_x(u)||p_{x,1}(u)) = \sum_u p_x(u) \log \frac{p_x(u)}{p_{x,1}(u)} \quad (27)$$

where l is the context image estimated by the segmentation algorithm. Considering that the ergodic theorem is the most fundamental principle in the detection and estimation theory, it is believed that when a good segmentation is achieved, the distance between the $p_x(u)$ and $p_{x,1}(u)$ should be minimized and

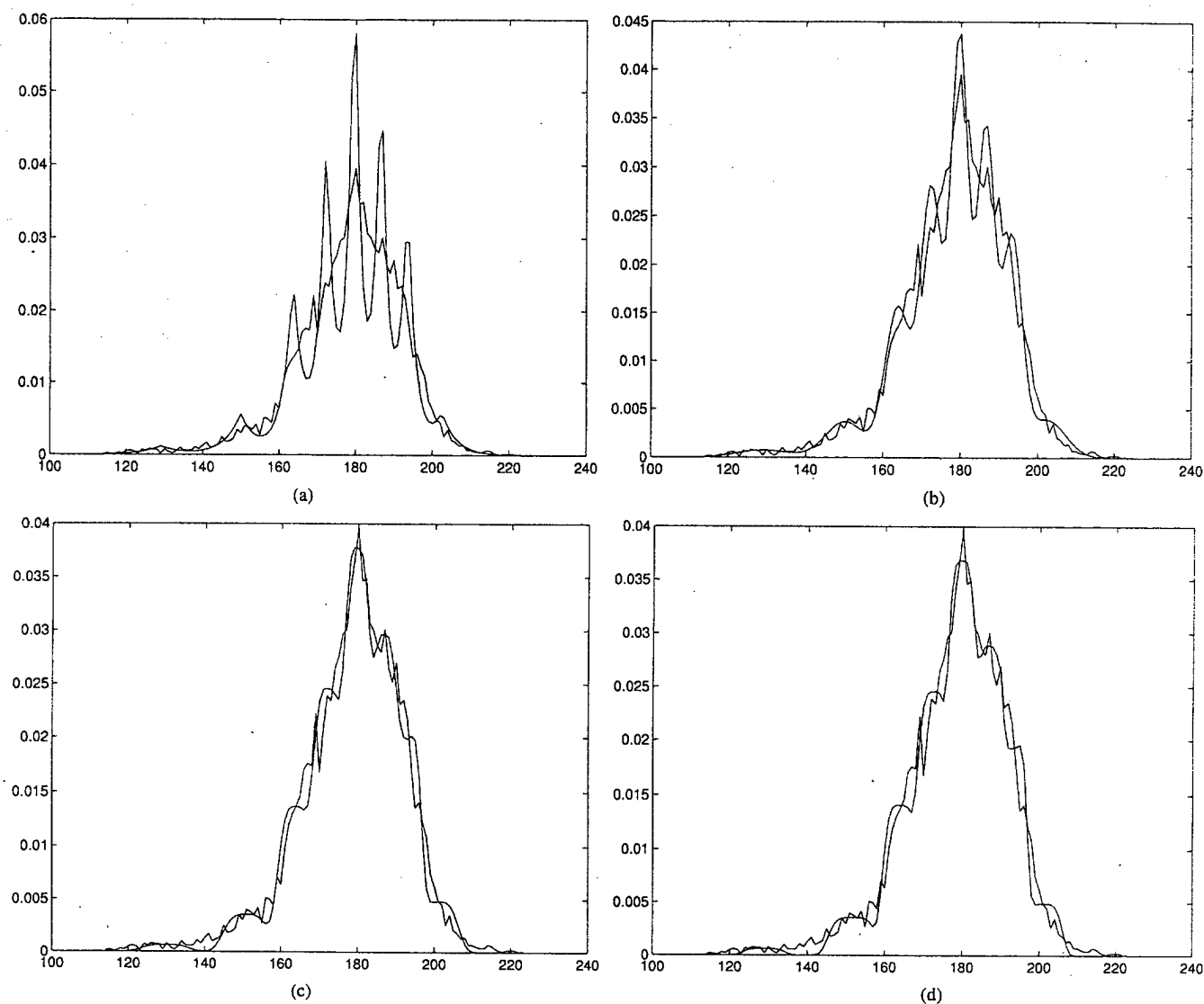


Fig. 7. Comparison of learning curves and histogram of the original mammogram with different α , $k = 8$. The optimal $\alpha = 3.0$. (a) $\alpha = 1.0$, GRE = 0.0783. (b) $\alpha = 2.0$, GRE = 0.0369. (c) $\alpha = 3.0$, GRE = 0.0251. (d) $\alpha = 4.0$, GRE = 0.0282.

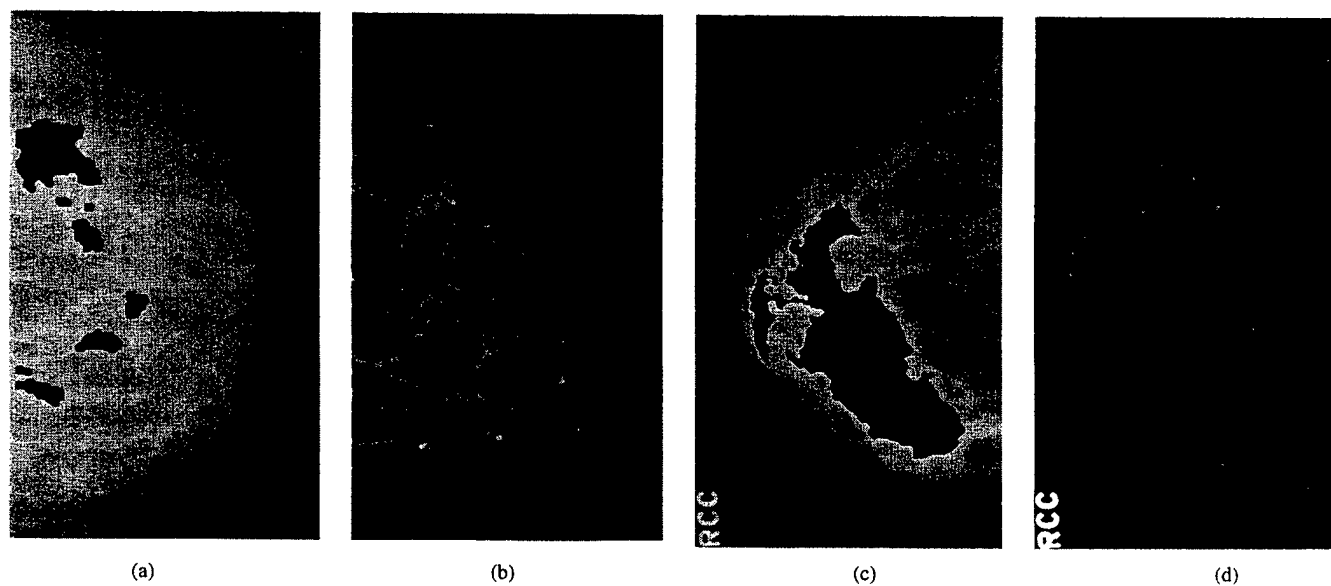


Fig. 8. Suspected mass segmentation results based on the original mammogram. (b) Result based on the enhanced mammogram, $K = 4$, $\alpha = 3.0$. (c) and (d) Results based on another original mammogram and its enhanced image.

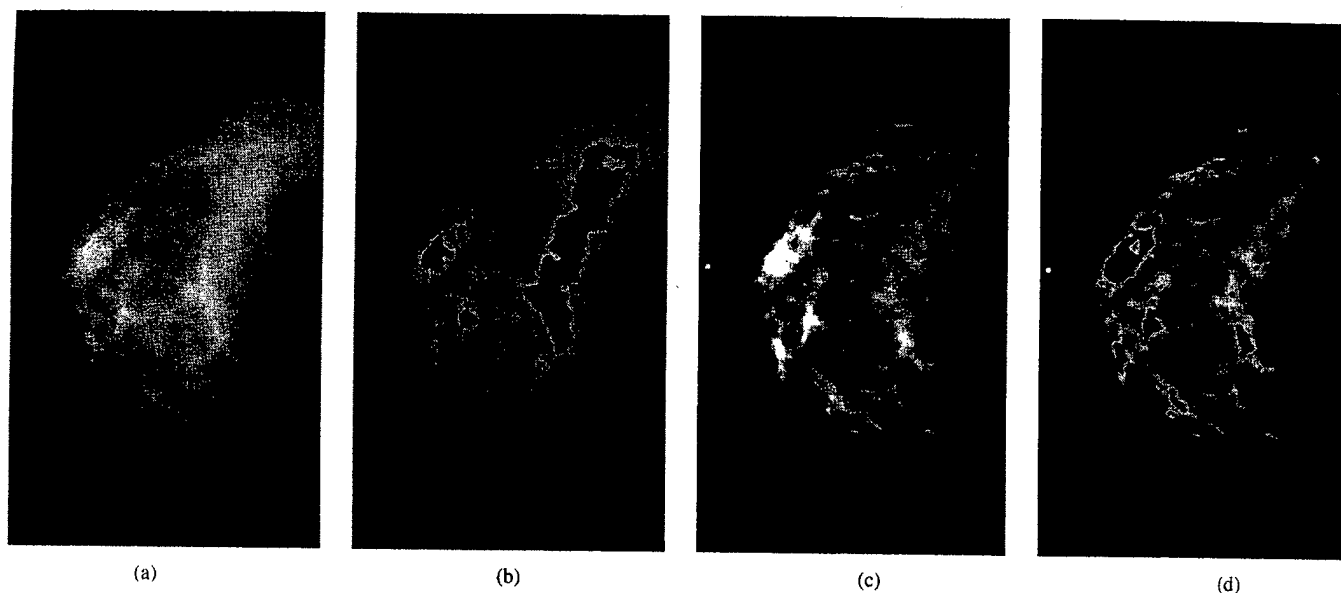


Fig. 9. Examples of normal mixed fatty and glandular mammogram. (a) Original mammogram. (b) Segmentation result based on the original mammogram. (c) Enhanced mammogram. (d) Result based on the enhanced mammogram, $k = 4$, $\alpha = 3.0$.

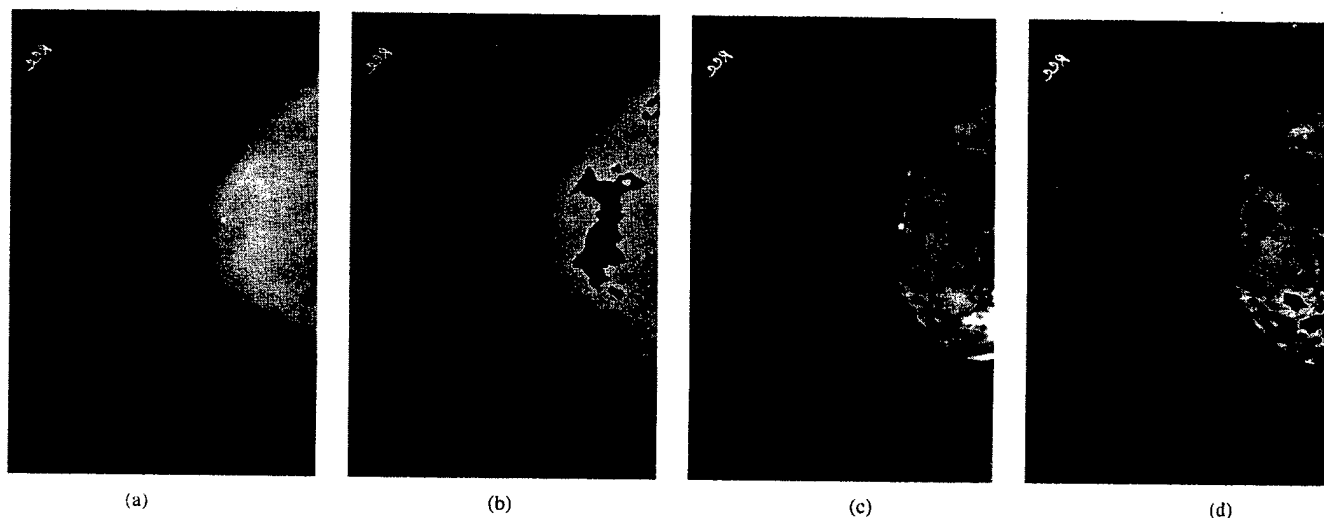


Fig. 10. Examples of normal dense mammogram. (a) Original mammogram. (b) Segmentation result based on the original mammogram. (c) Enhanced mammogram. (d) Result based on the enhanced mammogram, $k = 4$, $\alpha = 3.0$.

this measure links the image text and its sample averages. Our experience has suggested that this post-segmentation measure may be a suitable objective criterion for evaluating the quality of image segmentation in a fully unsupervised situation [22], [26]–[28]. Table V shows our evaluation data from three different segmentation methods when applied to the real images.

Performance of Combined Morphological Filtering and Model-Based Segmentation using a Larger Database: The proposed segmentation method was used to extract suspicious mass regions from the 200 testing mammograms. Without enhancement, a total of 1142 potential mass regions were isolated including 114 of the 186 true masses. With enhancement, a total of 3143 potential mass regions were extracted including 181 of the 186 true masses. The results demonstrated that more true masses were picked up after enhancement although more false cases were also included. The undetected areas mainly occurred at the lower intensity side of the shaded objects or obscured by fibroglandular tissues that, however, were extracted on morpho-

logical enhanced mammograms. In addition, when the margins of masses are ill defined, only parts of suspicious masses were extracted from the original mammograms. For the purpose of "lesion site selection," we believe that the sensitivity should be the sole criterion for the performance evaluation of the method. We have 181/186 versus 114/186. Our method is unsupervised and automatic and does not involve any detection effort at this moment. To our best knowledge, there is no objective criterion available for the evaluation of image enhancement performance before a detection effort is involved. We only claimed that the enhancement step is important and effective with respect to the purpose of "lesion site selection."

Fig. 8 demonstrates some segmentation results based on the original and enhanced mammograms. We compared the segmentation results based on the enhanced mammogram ($K = 4$, and $\alpha = 3.0$) with those based on the original mammogram ($K = 8$, and $\alpha = 3.0$) as shown in Fig. 8. Comparing the results in Fig. 8(b) with those in Fig. 8(a), we can see that after

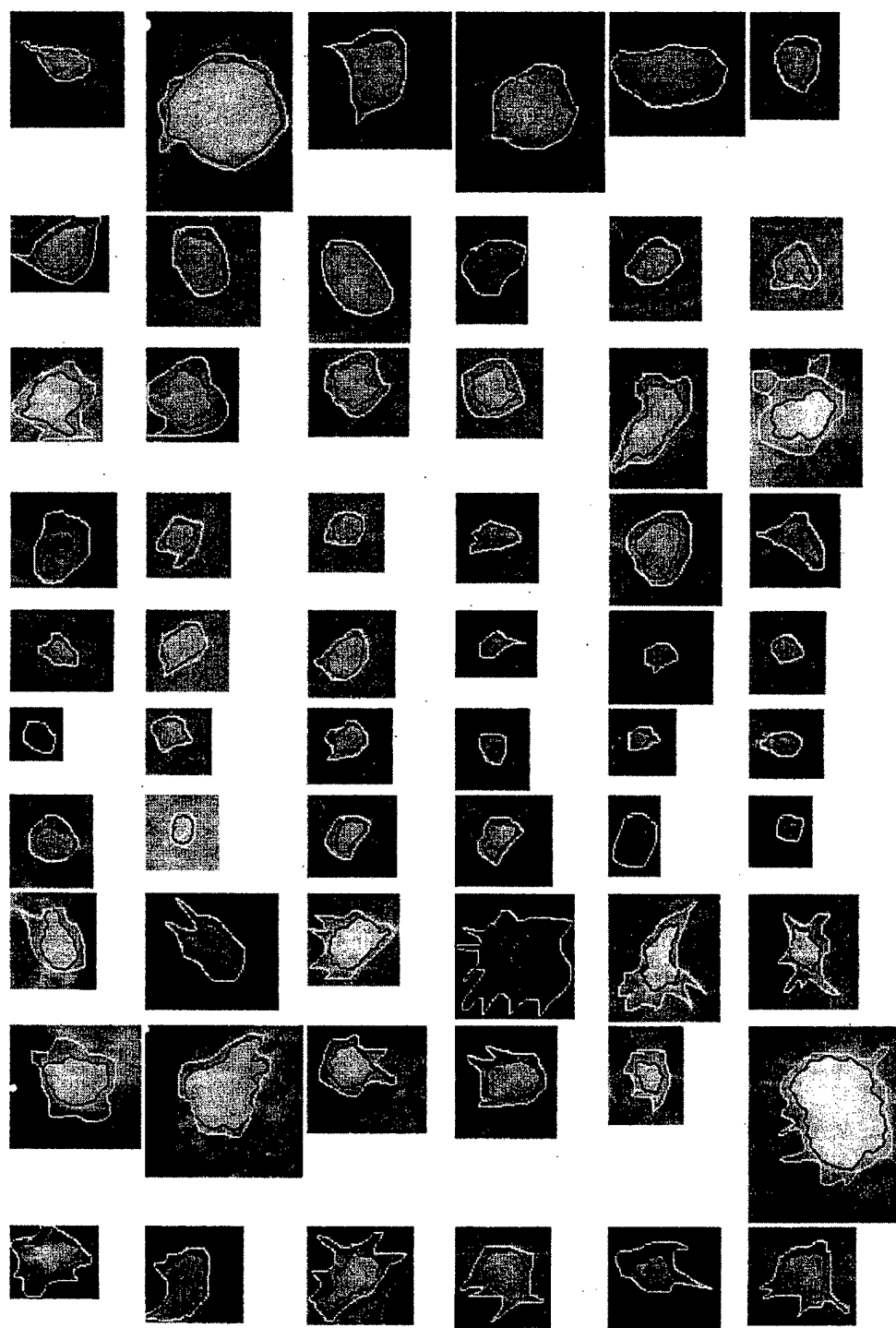


Fig. 11. Comparison results of segmentation based on the enhanced mammograms. Black outlines denote the computer-segmented result. White outlines denote the radiologist-segmented results.

enhancement, a more accurate region was detected for the suspected mass which has ill-defined margin. Getting an accurate suspected region is a crucial issue since geometric features are extracted based on suspected regions and these features are very important for further true mass detection. In addition, we observed that one suspected mass was missed in Fig. 8(a) but was detected in Fig. 8(b). As we have mentioned in Section I, none of the suspected masses should be missed in the segmentation step. Fig. 8(c) and (d) demonstrate the segmentation of a suspected

mass that lies in dense breast tissue. As shown in Fig. 8(c), the whole fibroglandular tissue area was segmented when based on the original mammogram. After enhancement, the suspected region was segmented exactly as shown in Fig. 8(d).

We have also included the segmentation results on the normal mammograms. Fig. 9 demonstrate the segmentation results based on the original and enhanced mixed fatty and glandular mammograms. Fig. 10 demonstrate the segmentation results based on the original and enhanced dense mammograms. We

would like to emphasize that the objective of this paper is to provide a segmentation technique which can enhance and extract potential mass site from the background so that the characterization of the related mass pattern can be accurately extracted in terms of focused feature selection and analysis. The method of course will produce many mass-like areas, but it will be a plausible outcome since the accurate description of nonmass cases characterized by mass-like sites will benefit the follow-on detection step where the performance of the classifier depends on an accurate separation of mass and nonmass in the featured spaces. The details will be described in [29].

For the purpose of evaluating the performance of the segmentation method, we used both simulated studies and expert visual inspection to validate the methods and results. The radiologist has concluded that the lesion characteristics after the proposed enhancement have been better displayed and all possible lesion areas have been successfully identified. In addition to the visual inspection, we have measured the overlap between the computer-segmented and the radiologist segmented mass regions to evaluate our method. Fig. 11 shows the comparison results of segmentation based on the enhanced mammograms. Fig. 11 includes 60 benign and malignant mass patches which were cut from the whole mammograms after the segmentation. The white outline was drawn by the radiologist while the black outline was produced by the computer and was superimposed upon the original image. As we can see from Fig. 11, for most of cases, the ratio of mutual overlap area of the radiologist segmented mass region and the computer-segmented mass region to the radiologist segmented mass area is large than 50%. In addition, even the poorest result picked the true lesion in the correct location and depicted the characteristics of the mass reasonably. It is important to understand that "lesion area segmentation" is not our objective, so there is no "best" or "worst" segmentation results. Our objective is "lesion site selection" with a possible highest sensitivity through a global unsupervised enhancement and segmentation scheme.

V. CONCLUSION

In this paper, we propose a combined method of using morphological operations, a FGGM modeling, and a CBRL to enhance and segment various breast tissue textures and suspicious mass lesions from mammographic images. This phase is a crucial step in mass detection for an improved CAD. We emphasized the importance of model selection which includes the selection of the number of image regions K and the selection of FGGM kernel shape controlled by α . The experimental results indicate that the suspected mass sites selection can be affected by different K and α . We proposed the EM algorithm together with the information theoretic criteria to determine the optimal K and α . With optimal K and α , the segmentation results can be significantly improved. We also showed that with the proposed pattern-dependent enhancement algorithm using morphological operations, the subtle masses can be segmented more accurately than those when the original image is used for extraction without enhancement. To summarize, the morphological filtering enhancement combined with the stochastic model-based segmentation is an effective way to extract mammographic suspicious

patterns of interest, and thereby may facilitate the overall performance of mammographic CAD of breast cancer.

ACKNOWLEDGMENT

The authors would like to thank Z. Gu of the Lombardi Cancer Center and I. Sesterhenn of the Armed Forces Institute of Pathology for their scientific input on the knowledge of cell oncology and pathology, and R. Shah MD, Director of Breast Imaging, BAMC for his evaluation of cases to our database.

REFERENCES

- [1] H. Kobatake, M. Murakami, H. Takeo, and S. Nawano, "Computerized detection of malignant tumors on digital mammograms," *IEEE Trans. Med. Imag.*, vol. 18, pp. 369–378, May 1999.
- [2] R. Zwiggelaar, T. C. Parr, J. E. Schumm, I. W. Hutt, C. J. Taylor, S. M. Astley, and C. R. M. Boggis, "Model-based detection of spiculated lesions in mammograms," *Med. Image Anal.*, vol. 3, no. 1, pp. 39–62, 1999.
- [3] M. A. Kupinski and M. L. Giger, "Automated seeded lesion segmentation on digital mammograms," *IEEE Trans. Med. Imag.*, vol. 17, pp. 510–517, Aug. 1998.
- [4] N. Karssemeijer and G. M. te Brake, "Detection of stellate distortions in mammogram," *IEEE Trans. Med. Imag.*, vol. 15, pp. 611–619, Oct 1996.
- [5] W. K. Zouras, M. L. Giger, P. Lu, D. E. Wolverton, C. J. Vyborny, and K. Doi, "Investigation of a temporal subtraction scheme for computerized detection of breast masses in mammograms," *Excerpta Medica*, vol. 1119, pp. 411–415, 1996.
- [6] N. Petrick, H. P. Chan, B. Sahiner, and D. Wei, "An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection," *IEEE Trans. Med. Imag.*, vol. 15, no. 1, pp. 59–67, 1996.
- [7] M. Sameti and R. K. Ward, "A fuzzy segmentation algorithm for mammogram partition," in *Digital Mammography*, ser. International Congress Series, K. Doi, Ed. Amsterdam, The Netherlands: Elsevier, 1996, pp. 471–474.
- [8] W. P. Kegelmeyer Jr., J. M. Pruneda, P. D. Bourland, A. Hillis, M. W. Riggs, and M. L. Nipper, "Computer-aided mammographic screening for spiculated lesions," *Radiology*, vol. 191, pp. 331–337, 1994.
- [9] F. F. Yin, M. L. Giger, C. J. Vyborny, K. Doi, and R. A. Schmidt, "Comparison of bilateral-subtraction and single-image processing techniques in the computerized detection of mammographic masses," *Investigat. Radiol.*, vol. 28, no. 6, pp. 473–481, 1993.
- [10] B. Zheng, Y. H. Chang, and D. Gur, "Computerized detection of masses in digitized mammograms using single-image segmentation and a multilayer topographic feature analysis," *Acad. Radiol.*, vol. 2, pp. 959–966, 1995.
- [11] H. D. Li, M. Kallergi, L. P. Clarke, V. K. Jain, and R. A. Clark, "Markov random field for tumor detection in digital mammography," *IEEE Trans. Med. Imag.*, vol. 14, pp. 565–576, Sept. 1995.
- [12] M. L. Giger, C. J. Vyborny, and R. A. Schmidt, "Computerized characterization of mammographic masses: Analysis of spiculation," *Cancer Lett.*, vol. 77, pp. 201–211, 1994.
- [13] T. K. Lau and W. F. Bischof, "Automated detection of breast tumors using the asymmetry approach," *Comput. Biomed. Res.*, vol. 24, no. 9, pp. 1501–1513, 1995.
- [14] M. J. Bianchi, A. Rios, and M. Kabuka, "An algorithm for detection of masses, skin contours, and enhancement of microcalcifications in mammograms," in *Proc., Symp. Computer Assisted Radiology*, Winston-Salem, NC, June 1994, pp. 57–64.
- [15] T. Lei and W. Sewchand, "Statistical approach to x-ray CT imaging and its application in image analysis—Part II: A new stochastic model-based image segmentation technique for x-ray CT image," *IEEE Trans. Med. Imag.*, vol. 11, pp. 62–69, Feb. 1992.
- [16] Y. Wang, T. Adali, and S.-C. B. Lo, "Automatic threshold selection using histogram quantization," *SPIE J. Biomedical Optics*, vol. 2, no. 2, pp. 211–217, April 1997.
- [17] J. Zhang and J. W. Modestino, "A model-fitting approach to cluster validation with application to stochastic model-based image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 1009–1017, Oct. 1990.

- [18] H. Li, K. J. R. Liu, Y. Wang, and S. C. Lo, "Morphological filtering and stochastic modeling-based segmentation of masses on mammographic images," in *Proc. IEEE Nuclear Science Symp. Medical Imaging Conf.*, 1996, pp. 1792–1796.
- [19] J. Serra, *Image Analysis and Mathematical Morphology*. London, U.K.: Academic, 1982.
- [20] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. Ser. B*, vol. 39, pp. 1–38, 1977.
- [21] Y. Wang, T. Adali, C. M. Lau, and S. Y. Kung, "Quantitative analysis of MR brain image sequences by adaptive self-organizing finite mixtures," *J. VLSI Signal Processing*, vol. 18, no. 3, pp. 219–240, 1998.
- [22] Y. Wang, T. Adali, S. Y. Kung, and Z. Szabo, "Quantification and segmentation of brain tissues from MR images: A probabilistic neural network approach," *IEEE Trans. Image Processing*, vol. 7, pp. 1165–1181, Aug. 1998.
- [23] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. 19, no. 6, pp. 716–723, 1974.
- [24] J. Rissanen, "Modeling by shortest data description," *Automat.*, vol. 14, pp. 465–471, 1978.
- [25] R. A. Hummel and S. W. Zucker, "On the foundations of relaxation labeling processes," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 5, pp. 267–286, Mar. 1983.
- [26] A. Hoover, G. J. Baptiste, X. Jiang, P. J. Flynn, H. Bunke, D. B. Goldgof, K. Bowyer, D. W. Eggert, A. Fitzgibbon, and R. B. Fisher, "An experimental comparison of range image segmentation algorithms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 673–688, July 1996.
- [27] Y. J. Zhang, "A survey on evaluation methods for image segmentation," *Pattern Recogn.*, vol. 29, no. 8, pp. 1335–1346, 1996.
- [28] A. M. Bensaid, L. O. Hall, J. C. Bezdek, L. P. Clarke, M. L. Silbiger, J. A. Arrington, and R. F. Murtagh, "Validity-guided clustering with applications to image segmentation," *IEEE Trans. Fuzzy Syst.*, vol. 4, pp. 112–123, May 1996.
- [29] H. Li, Y. Wang, K. J. R. Liu, S.-C. B. Lo, and M. T. Freedman, "Computerized Radiographic Mass Detection—Part II: Decision Support by Featured Database Visualization and Modular Neural Networks," *IEEE Trans. Med. Imag.*, vol. 20, no. 4, pp. 302–313, Apr. 2001.

Computerized Radiographic Mass Detection—Part II: Decision Support by Featured Database Visualization and Modular Neural Networks

Huai Li, Yue Wang, K. J. Ray Liu*, Shih-Chung B. Lo, and Matthew T. Freedman

Abstract—Based on the enhanced segmentation of suspicious mass areas, further development of computer-assisted mass detection may be decomposed into three distinctive machine learning tasks: 1) construction of the featured knowledge database; 2) mapping of the classified and/or unclassified data points in the database; and 3) development of an intelligent user interface. A decision support system may then be constructed as a complementary machine observer that should enhance the radiologists performance in mass detection. We adopt a mathematical feature extraction procedure to construct the featured knowledge database from all the suspicious mass sites localized by the enhanced segmentation. The optimal mapping of the data points is then obtained by learning the generalized normal mixtures and decision boundaries, where a is developed to carry out both soft and hard clustering. A visual explanation of the decision making is further invented as a decision support, based on an interactive visualization hierarchy through the probabilistic principal component projections of the knowledge database and the localized optimal displays of the retrieved raw data. A prototype system is developed and pilot tested to demonstrate the applicability of this framework to mammographic mass detection.

Index Terms—Feature extraction, knowledge database, mass detection, neural network, visual explanation.

I. INTRODUCTION

IN ORDER to improve mass lesion detection and classification in clinical screening and/or diagnosis of breast cancers, many sophisticated computer-assisted diagnosis (CAD) systems have been recently developed [1]–[10]. Although the clinical roles of the CAD systems may still be debatable, the fundamental role should be complementary to the radiologists'

Manuscript received February 3, 1997; revised January 9, 2001. This work was supported in part by the Department of Defense under Grants DAMD17-98-1-8045 and DAMD17-96-1-6254 through a subcontract from University of Michigan, Ann Arbor, and by the National Science Foundation (NSF) under NYI Award MIP-9457397. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was M. Giger. Asterisk indicates corresponding author.

H. Li is with the Electrical Engineering Department and Institute for Systems Research, University of Maryland at College Park, College Park, MD 20742 USA. He is also with the Department of Radiology, Georgetown University Medical Center, Washington, DC 20007 USA.

Y. Wang is with the Department of Electrical Engineering and Computer Science, The Catholic University of America, Washington, DC 20064 USA. He is also with the Department of Radiology, Georgetown University Medical Center, Washington, DC 20007 USA.

*K. J. Ray Liu is with the Electrical Engineering Department and Institute for Systems Research University of Maryland at College Park, College Park, MD 20742 USA (e-mail: kjrliu@eng.umd.edu).

S.-C. B. Lo and M. T. Freedman are with the Department of Radiology, Georgetown University Medical Center, Washington, DC 20007 USA.

Publisher Item Identifier S 0278-0062(01)02830-0.

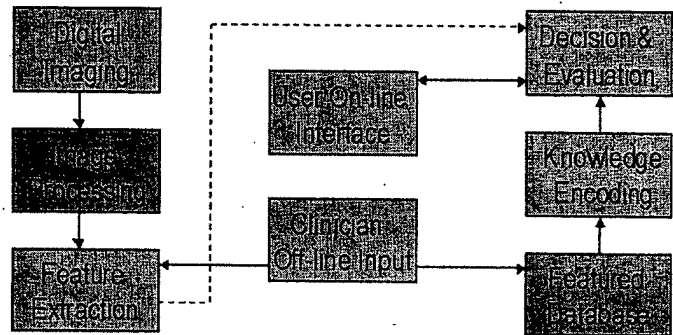


Fig. 1. Major components in CAD.

clinical duties, where the pathways of achieving ultimate performance enhancement taken by the machine observer and human observer may not necessarily be close. For example, CAD systems may attack the tasks that the radiologists cannot perform well or find difficult to perform. Because of generally larger size and complex appearance of masses, especially the existence of spicules in malignant lesions, as compared with microcalcifications, feature-based approaches are largely adopted in many CAD systems [1]–[4], [6], [7]. Kegelmeyer has first reported promising results for detecting spiculated tumors based on local edge characteristics and Laws texture features [7]. Zwiggelaar *et al.* developed a statistical model to describe and detect the abnormal pattern of linear structures of spiculated lesions [1]. Karssemeijer *et al.* [2] proposed to identify stellate distortions by using the orientation map of line-like structures. Petrick *et al.* presented to reduce the false positive detection by combining the breast tissue composition information [4]. Zhang *et al.* used the Hough spectrum to detect spiculated lesions [6].

Although many previously proposed approaches have led to impressive results [1]–[5], [7], several fundamental issues remain unresolved in the application of CAD systems. Fig. 1 shows a general block diagram of CAD systems. Previous research has demonstrated that: 1) breast cancer is missed on mammograms in part because the optical density and contrast of the cancer is not optimal for human observer; 2) computer-based detection appears to be more affected by different criteria than human perception; 3) the challenges and pathways to the human or machine observers may be quite different, and 4) decision making by the CAD systems are largely not transparent to the user. For example, the training cases contributing to the database are often selected by the human observer while the featured knowledge database is constructed through mathematical pathways of feature extraction. The mismatch

between the human supervised case selection in training and the machine dominant mass candidates selection in testing may exist. Second, the featured knowledge database is often high-dimensional with complex internal structures. Imposing a heuristically designed neural network for learning from the training data set may prevent a correct identification of the intrinsic data structure and an accurate estimation of the class boundaries. There may also exist the mismatch between the data structure and classifier architecture or between the class boundaries and decision boundaries. Furthermore, since the machine observer and human observer may not detect the same set of masses, the "black box" nature of most CAD systems to the clinical users will prevent a natural on-line integration of human intelligence and further upgrade of a CAD system. An interactive user interface should be considered to leverage the complementary roles of the CAD in the clinical practice.

As a step toward improving the performance of a CAD system, we have put considerable efforts to conduct various studies and develop reliable image enhancement and lesion selection techniques. The methods and results have been reported in [24], where the purposes of the research were to localize the potential mass sites and help accurate feature extraction. This paper addresses the further development of computer-assisted mass detection based on the 1) construction of the featured knowledge database; 2) mapping of the classified and/or unclassified data points in the database; and 3) development of an intelligent user interface (IUI). The clinical goal is to eliminate the false positive sites that correspond to normal dense tissues with *mass-like* appearances through featured discrimination. We adopt a mathematical feature extraction procedure to construct the featured knowledge database from all the suspicious mass sites localized by the enhanced segmentation. The optimal mapping of the data points is then obtained by learning the generalized normal mixtures and decision boundaries, where a probabilistic modular neural network (PMNN) is developed to carry out both soft and hard clustering. A visual explanation of the decision making is further invented as a decision support tool, based on an interactive visualization hierarchy through the probabilistic principal component projections of the knowledge database and the localized optimal displays of the retrieved raw data. The motivation of this work comes from the following considerations. First, though both human and machine observers use the same set of raw data in the diagnostic stage, the construction of the knowledge database for training machine classifiers and that accomplished by human brains are indeed different. Thus, the knowledge database should be established with both machine and expert organized representative cases. Second, a quantitative understanding of the knowledge database used by the machine observer should be acquired to logically compare and/or predict the performance of CAD systems with respect to the human observers without possible under- or over-estimation, and to optimize the feature extraction and design of the machine learner for best final performance. Finally, since the human and machine observers indeed take different learning and intelligence pathways, an IUI should be developed to visually (e.g., transparently) explain the entire internal decision making process of the CAD system to the human observer to enhance the clinical decision when facing either consistent or conflicting opinions.

The major differences between our work and the previous work [1]–[10] are as follows.

- 1) We construct a knowledge database by combining both expert and machine selected cases where the assignment of class memberships (e.g., mass and nonmass classes) is supervised by the radiologists or pathological report *after* all the cases are collected.
- 2) We impose a model identification procedure to determine the optimal number and kernel shape of the local clusters within each of the two classes in a high-dimensional feature space. The model is then estimated using the expectation-maximization (EM) algorithm and information theory.
- 3) We develop a PMNN, which is considered as a nonlinear classifier, to carry out the mapping function of the knowledge database. In the knowledge database, the decision likelihood boundaries and the class prior probabilities are determined in a separate fashion, and the structure of PMNN is optimized by adapting to the database structure.
- 4) We derive a probabilistic principal component projection scheme to reduce the dimensionality of the feature space for natural human perception. The scheme leads to a hierarchical visualization algorithm allowing the complete data set to be analyzed at the top level, with best separated clusters and subclusters of data points analyzed at deeper levels.

The framework of the proposed method for mass detection is illustrated in Fig. 2. A detailed description of this paper is organized as follows. In Section II, the procedure of the knowledge database construction is described. The data mapping process for decision making is presented in Section III. Section IV presents the design of the IUI for the CAD systems. Finally, major results and discussions are summarized in Section V.

II. KNOWLEDGE DATABASE CONSTRUCTION

Given the available information contained in the raw data of mass sites and in order to establish machine intelligence carried out by various machine observers, a knowledge database may be constructed in a multidimensional feature space. It should be emphasized however that the knowledge acquired by the human brain uses much more sophisticated processes than the artificial systems. Though feature extraction has been a key step in most pattern analysis tasks, the mathematical procedures are often done intuitively and heuristically. The general guidelines are:

- 1) *Discrimination*: Features of patterns in different classes should have significantly different values.
- 2) *Reliability*: Features should have similar values for the patterns of the same class.
- 3) *Independence*: Features should not be strongly correlated to each other.
- 4) *Optimality*: Some redundant features should be deleted. A small number of features is preferred for reducing the complexity of the classifier.

Many useful image features have been suggested previously by both image processing and pattern analysis communities [11]–[13]. These features can be divided into three categories, namely, intensity features, geometric features, and texture

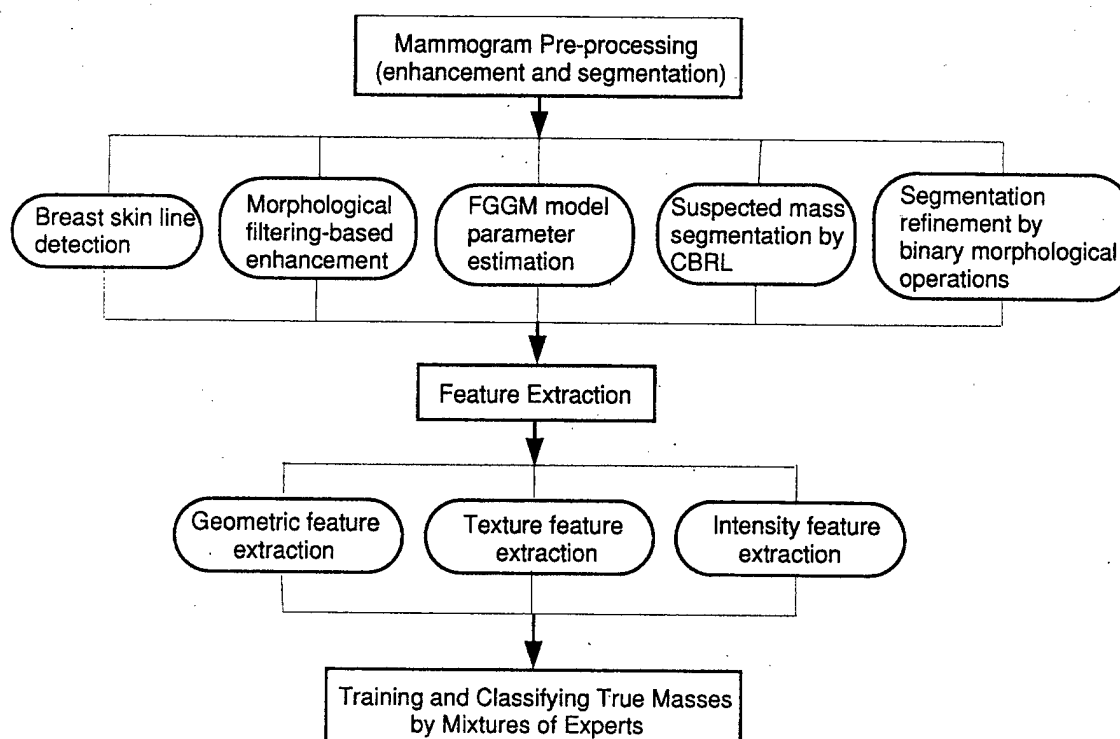


Fig. 2. The flow diagram of mass detection in digital mammograms.

features, whose values are calculated from the pixel matrices of the regions of interest (ROIs). Though these features are mathematically well defined, they may not be complete since they cannot capture all of the capable aspects of human perception nature. Thus, in this study, we have included several additional expert-suggested features to reflect the radiologists' experience. The typical features are summarized in Table I, where Fig. 3 shows the raw image of corresponding featured sites.

The joint histogram of the feature point distribution extracted from true and false mass regions are investigated, and the features that can better separate the true and false mass regions are selected for further study. Our experience has suggested that three features, i.e., the site area, two measured compactness (circularity), and difference entropy, were having better discrimination and reliability properties. Their definitions are given as follows.

1) *Compactness 1*

$$C_1 = \frac{A_1}{A} \quad (1)$$

where A is the area of the actual suspected region, and A_1 is the area of the overlapped region of A and the effective circle A_c , which is defined as the circle whose area is equal to A and is centered about the corresponding centroid of A .

2) *Compactness 2*

$$C_2 = \frac{P}{4\pi A} \quad (2)$$

where P is the boundary perimeter, and A is the area of region.

TABLE I
THE SUMMARY OF MATHEMATICAL FEATURES

Feature Sub-Space	Features
A. Intensity Features	1. contrast measure of ROIs; 2. standard derivation inside ROIs; 3. mean gradient of ROIs boundary
B. Geometric Features	1. area measure; 2. circularity measure; 3. deviation of the normalized radial length; 4. boundary roughness;
C. Texture Features	1. energy measure; 2. correlation of co-occurrence matrix; 3. inertia of co-occurrence matrix; 4. entropy of co-occurrence matrix; 5. inverse difference moment; 6. sum average; 7. sum entropy; 8. difference entropy; 9. fractal dimension of surface of ROI;

3) *Difference Entropy*

$$DH_{d,\theta} = - \sum_{k=0}^{L-1} p_{x-y}(k) \log p_{x-y}(k) \quad (3)$$

where

$$p_{x-y}(k) = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p_{d,\theta}(i, j), \quad |i - j| = k. \quad (4)$$

Several important observations are worth reiteration:

- 1) The knowledge database that will be used by the CAD system are constructed from the cases selected by both lesion localization procedure and human expert's experience. This joint set provides more complete knowledge to

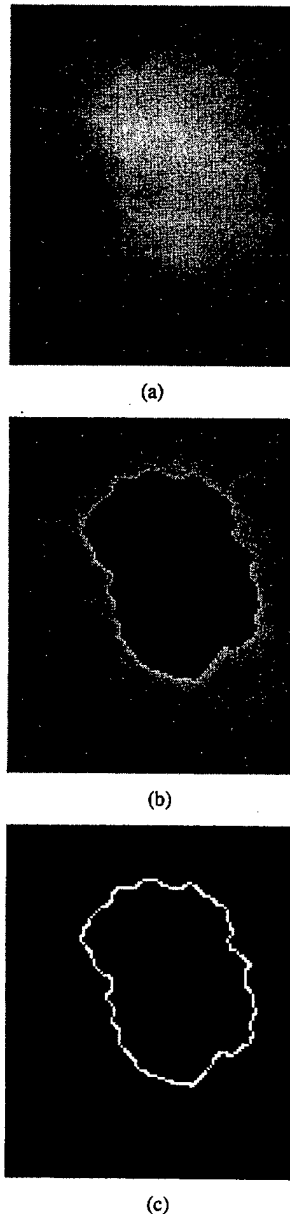


Fig. 3. One example of mass segmentation and boundary extraction. (a) Mass patch; (b) segmentation; (c) boundary extraction.

the machine observer. In particular, during the interactive decision making, CAD system can still provide opinion when the cases are missed by the localization procedure but presented to the system by the radiologists.

- 2) The knowledge database is defined quantitatively in a high dimensional feature space. It provides not only the knowledge for training the machine observer, but also an objective base for evaluating the quality of feature extraction or network's learning capability, and the on-line visual explanation possibility.
- 3) The assignment of the cases' class memberships (e.g., mass and nonmass classes) is supervised by the radiologists or pathological reports. A complete knowledge database includes three subsets: raw data of mass-like sites, corresponding feature points, and class membership labels.

III. DATA MAPPING FOR DECISION MAKING

The decision making support by a CAD system addresses the problem of mapping a knowledge database, given a finite set of data examples. The mapping function can therefore be interpreted as a quantitative representation of the knowledge about the mass lesions contained in the database [14]. Instead of mapping the whole data set using a single complex network, it is more practical to design a set of simple class subnets with local mixture clusters, each one of which represents a specific region of the knowledge space. Inspired by the principle of *divide-and-conquer* in applied statistics, PMNN has become increasingly popular in machine learning research [14], [15], [19]–[22]. In this section, we present its applications to the problem of mapping from databases in mass detection, with a constructive criterion for designing the network architecture and the learning algorithm that are governed by information theory [25].

A. Statistical Modeling

The quantitative mapping of a database may be decomposed into three distinctive learning tasks: the detection of the structure of each class model with local mixture clusters; the estimation of the data distributions for each induced cluster inside each class; and the classification of the data into classes that realizes the data memberships. Recently, there has been considerable success in using finite mixture distributions data mapping [15], [17], [18], [20]. Assume that the data points \vec{x}_i in a multidimensional database come from M classes $\{\vec{\omega}_1, \dots, \vec{\omega}_r, \dots, \vec{\omega}_M\}$, and each class contains K_r clusters $\{\vec{\theta}_1, \dots, \vec{\theta}_k, \dots, \vec{\theta}_{K_r}\}$, where $\vec{\omega}_r$ is the model parameter vector of class r , and $\vec{\theta}_k$ is the kernel parameter vector of cluster k within class r . The class conditional probability measure for any data point inside the class r , i.e., the standard finite mixture distribution (SFMD), can be obtained as a sum of the following general form:

$$f(\vec{u}|\vec{\omega}_r) = \sum_{k=1}^{K_r} \pi_k g(\vec{u}|\vec{\theta}_k) \quad (5)$$

where $\pi_k = P(\vec{\theta}_k|\vec{\omega}_r)$ with a summation equal to one, and $g(\vec{u}|\vec{\theta}_k)$ is the kernel function of the local cluster distribution. For the model of global class distributions, we denote the Bayesian prior for each class by $P(\vec{\omega}_r)$. Then the sufficient statistics according to the Bayes' rule, are the posterior probability $P(\vec{\omega}_r|\vec{x}_i)$ given a particular observation \vec{x}_i

$$P(\vec{\omega}_r|\vec{x}_i) = \frac{P(\vec{\omega}_r)f(\vec{x}_i|\vec{\omega}_r)}{p(\vec{x}_i)} \quad (6)$$

where $p(\vec{x}_i) = \sum_{r=1}^M P(\vec{\omega}_r)f(\vec{x}_i|\vec{\omega}_r)$.

B. Class Distribution Learning

Class distribution learning addresses the combined estimation of regional parameters ($\pi_k, \vec{\theta}_k$) and detection of the structural parameter K_r and the kernel shape of $g(\cdot)$ in (5) based on the observations \mathbf{x}_r . One natural criterion used for learning the optimal parameter values is to minimize the distance between the SFMD, denoted by $f_r(\vec{u})$, and the class data histogram, denoted by $f_{\mathbf{x}_r}(\vec{u})$ [17]. In this paper, we use relative entropy (Kullback–Leibler distance), suggested by information theory

[25], as the distance measure (for simplicity we use $f_r(\vec{u})$ to denote $f(\vec{u}|\vec{\omega}_r)$ in our formulation), given by

$$D(f_{x_r}||f_r) = \sum_{\vec{u}} f_{x_r}(\vec{u}) \log \frac{f_{x_r}(\vec{u})}{f(\vec{u}|\vec{\omega}_r)}. \quad (7)$$

We have previously shown that when relative entropy is used as a distance measure, the distance minimization method is equivalent to the soft-split classification-based method under the criterion of maximum likelihood (ML) [23].

Another important issue concerning unsupervised distribution learning is the detection of the structural parameters of the class distribution, called model selection [15]. The objective here is to propose a systematic strategy for determining the optimal number and kernel shape of local clusters, when the prior knowledge is not available. This is indeed the case when the structure of the mass lesion patterns for a particular type of cancer may be arbitrarily complex, so correct identification of the database structure is very important. Thus, it will be desirable to have a neural network structure that is adaptive, in the sense that the number and kernel shape of local clusters are not fixed beforehand. In this paper, we applied two popular information theoretic criteria, i.e., the Akaike information criterion and minimum description length to guide the model selection procedure [24].

As the counterpart for adaptive model selection, there are many numerical techniques to perform ML estimation of cluster parameters [17]. For example, EM algorithm first calculates the posterior Bayesian probabilities of the data through the observations and the current parameter estimates (*E*-step) and then updates parameter estimates using generalized mean ergodic theorems (*M*-step). The procedure cycles back and forth between these two steps. The successive iterations increase the likelihood of the model parameters. The scheme provides winner-takes-in probability (Bayesian "soft") splits of the data, hence allowing the data to contribute simultaneously to multiple clusters. For the sake of simplicity, we assume the kernel shape of local clusters to be a multidimensional Gaussian with mean $\vec{\mu}_{kr}$ and variance Γ_{kr} . We summarize the EM algorithm as follows.

- 1) **E-Step:** for training sample $\vec{x}^{(t)}$, $t = 1, \dots, N$, compute the probabilistic membership

$$h_{kr}^{(m)}(t) = \frac{\pi_{kr}^{(m)} p_k^{(m)}(\vec{x}^{(t)}|\vec{\omega}_r)}{\sum_{k=1}^{K_r} \pi_{kr}^{(m)} p_k^{(m)}(\vec{x}^{(t)}|\vec{\omega}_r)}. \quad (8)$$

- 2) **M-Step:** compute the updated parameter estimates

$$\pi_{kr}^{(m+1)} = \frac{1}{N} \sum_{t=1}^N h_{kr}^{(m)}(t) \quad (9)$$

$$\vec{\mu}_{kr}^{(m+1)} = \frac{1}{N\pi_{kr}^{(m+1)}} \sum_{t=1}^N h_{kr}^{(m)}(t) \vec{x}^{(t)} \quad (10)$$

$$\Gamma_{kr}^{(m+1)} = \frac{1}{N\pi_{kr}^{(m+1)}} \sum_{t=1}^N h_{kr}^{(m)}(t) \left[\vec{x}^{(t)} - \vec{\mu}_{kr}^{(m+1)} \right] \times \left[\vec{x}^{(t)} - \vec{\mu}_{kr}^{(m+1)} \right]^T. \quad (11)$$

C. Decision Boundary Learning

The objective of data classification is to realize the class membership l_i for each data points based on the observation \vec{x}_i and the class statistics $\{P(\vec{\omega}_r), f(\vec{u}|\vec{\omega}_r)\}$. It is well known that the optimal data classifier is the Bayes classifier since it can achieve the minimum rate of classification error [26]. Measuring the average classification error by the mean squared error E , many previous researchers have shown that minimizing E by adjusting the parameters of class statistics is equivalent to directly approximating the posterior class probabilities when dealing with the two class problem [13], [26]. In general, for the multiple class problem the optimal Bayes classifier (minimum average error) classifies input patterns based on their posterior probabilities: input \vec{x}_i is classified to class $\vec{\omega}_r$ if

$$P(\vec{\omega}_r|\vec{x}_i) > P(\vec{\omega}_j|\vec{x}_i) \quad (12)$$

for all $j \neq r$. It should be noted that in the formulation of classifier design, the optimal criterion used for the future data classification has been intuitively and directly applied to the learning of class statistics from the training data set.

Direct learning of posterior probability is a complex task. Great effort has been made in designing the classifier as an estimator of the posterior class probability [19]. By closely investigating the global class distribution modeling, we found that the classifier design for data classification can be dramatically simplified at the learning stage. Revisit (6), since the class prior probability $P(\vec{\omega}_r)$ is a known parameter when a supervised learning is applied, the posterior class probability $P(\vec{\omega}_r|\vec{x}_i)$ can be obtained without any further effort. Thus, by conditioning $P(\vec{\omega}_r)$, the problem is formulated as a supervised classification learning of the class conditional likelihood density $f(\vec{u}|\vec{\omega}_r)$. Thus, an efficient supervised algorithm to learn the class conditional likelihood densities called the "decision-based learning" [21] is adopted in this paper. The decision-based learning algorithm uses the *misclassified* data to adjust the density functions $f(\vec{u}|\vec{\omega}_r)$, which are initially obtained using the unsupervised learning scheme described previously, so that the minimum classification error can be achieved. Define the r th class discriminant function $\phi_r(\vec{x}_i, \mathbf{w})$ to be $P(\vec{\omega}_r)f(\vec{x}_i|\vec{\omega}_r)$. Given a set of training patterns $\mathbf{X} = \{\vec{x}_i; i = 1, 2, \dots, M\}$. The set \mathbf{X} is further divided into the "positive training set" $\mathbf{X}^+ = \{\vec{x}_i; \vec{x}_i \in \vec{\omega}_r, i = 1, 2, \dots, N\}$ and the "negative training set" $\mathbf{X}^- = \{\vec{x}_i; \vec{x}_i \notin \vec{\omega}_r, i = N+1, N+2, \dots, M\}$. If the misclassified training pattern is from positive training set, reinforced learning will be applied. If the training pattern belongs to the negative training set, we anti-reinforce the learning, i.e., pull the kernels away from the problematic regions. The boundary refinement is summarized as follows:

Reinforced

$$\text{Learning: } \mathbf{w}^{(j+1)} = \mathbf{w}^{(j)} + \eta l'(d(t)) \nabla \phi(\mathbf{x}(t), \mathbf{w})$$

Antireinforced

$$\text{Learning: } \mathbf{w}^{(j+1)} = \mathbf{w}^{(j)} - \eta l'(d(t)) \nabla \phi(\mathbf{x}(t), \mathbf{w}) \quad (13)$$

PMNN is a probabilistic modular network designed especially for data classification where a Bayesian decomposition of the learning process provides a unique opportunity to optimize

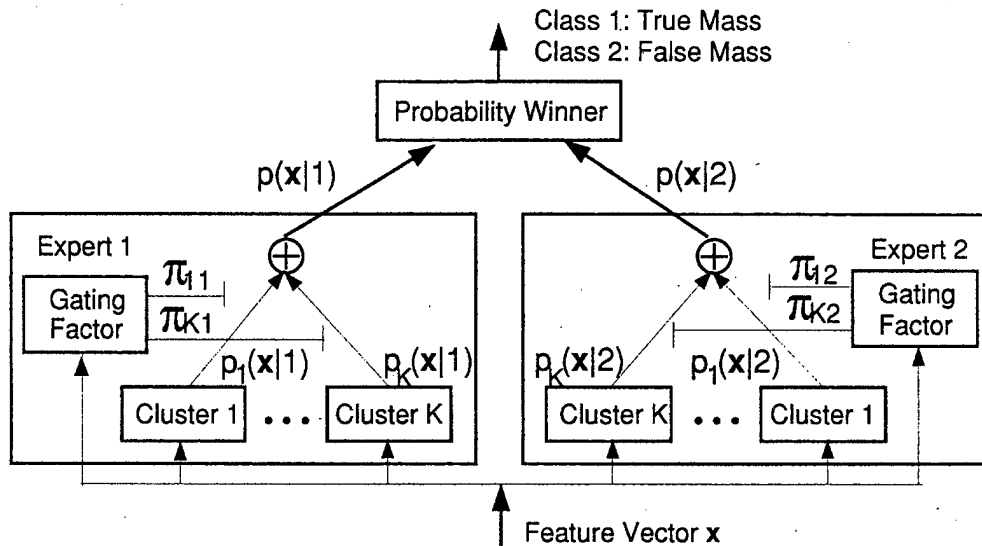


Fig. 4. The structure of the PMNN.

the structure of training scheme [14], [22]. Since the information about class population is, in general, physically uncorrelated with the conditional features about the individual class, a decoupled two-step training, in terms of both network structure and learning rule, makes much more sense than that in the conventional posterior-type neural networks, i.e., the conditional likelihood of each class and the class Bayesian prior should be adjusted separately in the classification spaces. Thus, PMNN consists of several disjoint subnets and a winner-takes-all network. The subnet outputs of the PMNN are designed to model the likelihood functions (likelihood-type network) which are first estimated from equally presented class samples, and the final decision boundaries are determined simply weighting the likelihood by the class populations. For a M -classification problem, PMNN contains M different class subnets, each of which represents one data class in the database. Within each subnet, several neurons (or clusters) are applied in order to handle problems which have complicated decision boundaries. The outputs of class subnets are fed into a winner-take-all network. The winner-take-all network categorizes the input pattern to the data class whose subnet produces the highest output value.

The structure of the PMNN used in this study is shown in Fig. 4. The PMNN consists of two subnets. Within each subnet, there are several neurons (or clusters). The outputs of class subnets are fed into a probability winner processor, which categorizes the input pattern to the data class whose subnet produces the highest probability value. The training scheme of the PMNN is based on the unsupervised learning. Each subnet is trained individually, and no mutual information across the classes may be utilized. In our study, one modular expert is trained to detect true masses, and the other is trained to detect false masses. After training, the feature vectors extracted from ROIsub are entered to this network to classify true or false masses. In both training and testing processes, we assume that the feature vectors \vec{x}_i in class r ($r = 1, \dots, M$) is a mixture of multidimensional Gaussian distributions, i.e.,

$$f(\vec{x}_i|\vec{\omega}_r) = \sum_{k=1}^{K_r} \pi_{kr} p_k(\vec{x}_i|\vec{\omega}_r) \quad (14)$$

where $\sum_{k=1}^{K_r} \pi_{kr} = 1$ and $p_k(\vec{\omega}_r) = N(\vec{\mu}_{kr}, \Gamma_{kr})$ is a multi-dimensional Gaussian distribution within cluster k of class r .

IV. INTERACTIVE VISUAL EXPLANATION

In order to improve the utility of the CAD systems in clinical practice, an IUI is highly desired. Different from many previously proposed approaches, we have organized our database from both mathematical-localized and radiologist-selected mass-like cases, and formed the featured knowledge database based on both mathematical-based and radiologist-selected image features. This off-line effort should enhance the performance of the machine observer through better quality of training set and optimal design of neural network architecture. Our experience has suggested, however, that further improvement of CAD systems requires on-line natural integration of human intelligence with the computer's output, since human perception has and can play an important role in the clinical decision making. In this research, we have pilot developed an IUI where the major functions include: 1) interactive visual explanation of the CAD decision making process; 2) on-line retrieval of the optimally displayed raw data and/or similar cases; and 3) supervised upgrade of the knowledge database by radiologist-driven input of the "unseen" and/or "typical" cases. Our preliminary studies have shown that the visual presentation of both raw data and CAD results to radiologists may provide visual cues for improved decision making.

As a step toward understanding the complex information from data and relationships, structural and discriminative knowledge reveals insight that may prove useful in data mining. Hierarchical minimax entropy modeling and probabilistic principal component projection are proposed for data explanation, which is both statistically principled and visually effective at revealing all of the interesting aspects of the data set. The methods involve multiple use of standard finite normal mixture models and probabilistic principal component projections. The strategy is that the top-level model and projection should explain the entire data set, best revealing the presence of clusters and relationships, while lower-level models and

projections should display internal structure within individual clusters, such as the presence of subclusters and attribute trends, which might not be apparent in the higher-level models and projections. With many complementary mixture models and visualization projections, each level will be relatively simple while the complete hierarchy maintains overall flexibility yet still conveys considerable structural information. In particular, a probabilistic principal component neural network is developed to generate optimal projections, leading to a hierarchical visualization algorithm. This algorithm allows the complete data set to be analyzed at the top level, with best separated subclusters of data points analyzed at deeper levels.

Research evidence suggests that for analysis of complex and high-dimensional data sets, structure decomposition and dimensionality reduction are the natural strategies in which the model-based approach and visual explanation have proven to be powerful and widely-applicable [27]. However, there is a trade-off between maximizing (structure decomposition) and minimizing (dimensionality reduction) the entropy of the system. In this research, a minimax entropy approach is adopted through the use of progressive model identification and principal component projection. The complete visual explanation hierarchy is generated by performing principal projection (dimensionality reduction) and model identification (structure decomposition) in two iterative steps using information theoretic criteria, EM algorithm, and probabilistic principal component analysis (PCA). Hierarchical probabilistic principal component visualization involves: 1) evaluation of posterior probabilities for mixture data set; 2) estimation of multiple principal component axes from probabilistic data set; and 3) generation of a complete hierarchy of visual projections.

Suppose the data space is d -dimensional with coordinates y_1, \dots, y_d and the data set consists of a set of d -dimensional vectors $\{t_i\}$ where $i = 1, \dots, N$. Now consider a three-dimensional (3-D) latent space $\mathbf{x} = (x_1, x_2, x_3)^T$ together with a linear function which maps the latent space to the data space by $\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{b}$ where \mathbf{W} is a $d \times 3$ matrix and \mathbf{b} is a d -dimensional mean vector. If we introduce a probability distribution $p(\mathbf{x})$ over the latent space given by a Gaussian estimated from the latent variables $\{x_i\}$, then a similar full-dimensional Gaussian distribution in data space can be defined by convolving this distribution with a general diagonal Gaussian conditional probability distribution $p(t|\mathbf{x}, \Lambda_d)$ in data space where Λ_d is the covariance matrix, resulting in a final form of

$$p(t) = \int p(t|\mathbf{x})p(\mathbf{x})d\mathbf{x} \quad (15)$$

where the log likelihood function for this model is given by $L = \sum_i \log p(t_i)$. Suppose \mathbf{W} is determined by the PCA, ML can be used to fit the model to the data and hence determine values for the parameters \mathbf{b} and Λ_d [27]. Using a soft clustering of the data set and multiple PCA sub corresponding to the clusters, a mixture of latent models takes the form of $p(t) = \sum_{k=1}^{K_0} \pi_k p(t|k)$ where K_0 is the number of components in the mixture, and the parameters π_k are the prior probabilities corresponding to the components $p(t|k)$. Each component is an independent latent model with PCA projection \mathbf{W}_k and parameters \mathbf{b}_k and Λ_{dk} . This procedure can be further extended to a hierarchical mixture model formulated by $p(t) = \sum_{k=1}^{K_0} \pi_k \sum_j \pi_{j|k} p(t|k, j)$

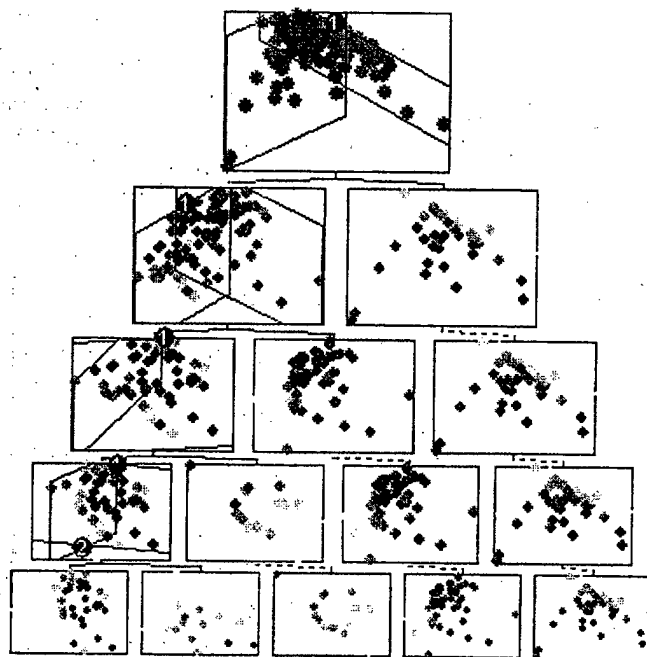


Fig. 5. The hierarchical view of computed features for mass and nonmass samples (Database A, see Table II).

where $p(t|k, j)$ again represent independent latent models [27]. With a soft partitioning of the data set via EM algorithm, data points will effectively belong to more than one cluster at any given level. This step is automatically available in our approach since the estimation of parent latent model involves the calculation of posterior probabilities denoted by z_{ik} . Thus, the effective input values are $z_{ik}\mathbf{x}_i$ for an independent visualization space k , corresponding to the visualization space k in the hierarchy. It should be emphasized that *probabilistic* means both neural network based learning and posterior probability weighted inputs. Further projections can again be performed by using the effective input values $z_{ik}z_{j|k}t_i$ for the visualization subspace j . Fig. 5 shows the hierarchical view of computed features for mass and nonmass samples. In Fig. 5, a hierarchical visualization view of a high dimensional feature data set was generated using hierarchical data visualization algorithm. One hundred and 25 real cases were involved, among them 75 are mass sites, 50 are nonmass sites. Nine features were computed on 125 cases. The dimension of the resulted feature data set became 125×9 (Database A, see Table II). Hierarchical visualization tool enables the visualization of high dimensional data set through dimension reduction and data modeling so that data distribution features of the data set can be well recognized. For instance, the clusters and subclusters of mass and nonmass data points and the boundaries of the clusters can be revealed for further research purpose.

In the use of a hierarchical minimax entropy mixture model, an interactive visualization environment is required to enable a flexible computerized experiment such that a human-database interaction can be performed effectively. We have developed an interactive environment for visualizing five-dimensional (5-D) data sets, based on state-of-the-art computer graphics toolkits such as object-oriented OpenGL and OpenInventor. With a sophisticated set of various kinds of simulated lights, color

TABLE II
THE SUMMARY OF EXPERIMENTAL DATABASES

Database	Descriptions
A	Nine features extracted from 75 mass sites and 50 non-mass sites. Used for visualizing hierarchically projected high dimensional feature space. Result is presented in Figure 5.
B	A simulated two-dimensional feature space. Used to show the effect of model selection on decision boundary estimation. Result is shown in Figure 6.
C	ORL standard database. Used to show the improvement of PMNN with decision-based learning. Result is discussed in the text.
D	The training data set consisting of 50 mammograms, with 50 true mass sites and 50 false mass sites. Three most discriminatory features are extracted. Used for both PMNN training and visualization. Result is given in Figure 7.
E	The testing data set consisting of 46 mammograms, with 23 normal cases and 23 biopsy proven mass cases with each of them having at least one true mass site. Three most discriminatory features, the same as database D, are extracted. Used to test the overall performance of our CAD system prototype where the mass candidates were selected using the method reported in Part I, automatically. Result is shown in Figure 8 and also discussed in the text.

texturing editors, and 3-D manipulator and viewers (we have integrated 3-D mouse and stereo glass units into our existing system), our system allows one to examine the volumetric data sets with any viewpoint and dynamically walk through its internal structures to better understand the spatial relationships among clusters and decision surfaces present. One of the most important features in our approach is to attach the decision surface to the 3-D probability cloud in support of decision making, and to link each data point in the visualization space to its raw data so that the user can on-line retrieve the corresponding raw data such as an original image for interim decision making.

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, we present the experimental results using the information theoretic criteria and PMNNs to generate the mapping function of the featured database, and the preliminary results using the hierarchical minimax entropy projections to conduct visual explanation of the decision making. For the validation of the database mapping using the proposed algorithms, global relative entropy (GRE) value between the (SFMD) and the joint histogram is used as an objective measure to evaluate the fitness of the mapping function. A summary of the databases we used in our study is presented in Table II.

As we have discussed in Sections III and IV, model selection is the first and a very important learning task in mapping a database and the objective of the procedure is to determine both the number and the kernel shape of local clusters in each class. This procedure is used not only in the data mapping for decision making but also in the structure decomposition for hierarchical visual explanation. Our experience has suggested that an incorrect model selection will affect the performance of data-classification based decision making. For the sake of simplicity, we discuss this conclusion in the following 2-D example. Let us form a simulated featured database with two major features that well characterize the two targeted classes, as it shown in Fig. 6 (Database B, see Table II). The ground truth is that class 1 contains only one local cluster while class 2 contains two local clusters. With a model selection procedure

using the proposed criteria, the intrinsic data structure was correctly identified. According to the principle of designing the optimal structure of PMNN and visual explanation hierarchy, the result of these criteria also determines the most appropriate number of mixture components in the corresponding PMNN and projected cluster decomposition. Two PMNN with different architecture orders were designed and trained to determine the classification boundaries between the two classes. The classification results are shown in Fig. 6(a) and (b). The result in Fig. 6(a) is with the right cluster number in Class 2, while the result in Fig. 6(b) is with the wrong cluster number in Class 2. From this simple experiment, we have shown that the decision boundary with the right cluster number may be much more accurate than that with heuristically determined cluster number, since the decision boundary between class 1 and class 2 will be determined by four cross points in the first case while in the second case the decision boundary will be determined by only two cross points. It should be emphasized that the error of data classification is theoretically controlled by the accuracy in estimating the decision boundaries between classes, and the quality of the boundary estimates is indeed dependent upon the correct structure of the class likelihood function.

As we have discussed before, although the knowledge database contains both machine-localized and human-selected cases, in clinical settings "unseen" and/or subtle cases contribute the major false positives. We have also pilot tested the PMNN method to the so-called " $M + 1$ classes" problem, in which the disease pattern under testing could be either from one of the M classes, or from some other unknown classes (the "unknown" class or the "intruder" class). Note that the unknown class probability is often very hard to estimate because of the lack of sufficient training samples (for example, in the mass detection problem, the unknown classes include the ROIs over the normal tissues). In our experiment, PMNN uses different decision rule from that of the " M classes" problem: pattern \vec{x}_i belongs to class r if both of the following conditions are true: a) $\phi(\vec{\omega}_r, \vec{x}_i) > \phi(\vec{\omega}_j, \vec{x}_i), \forall j \neq r$, and b) $\phi(\vec{\omega}_r, \vec{x}_i) > T$. T is a threshold obtained by decision-based

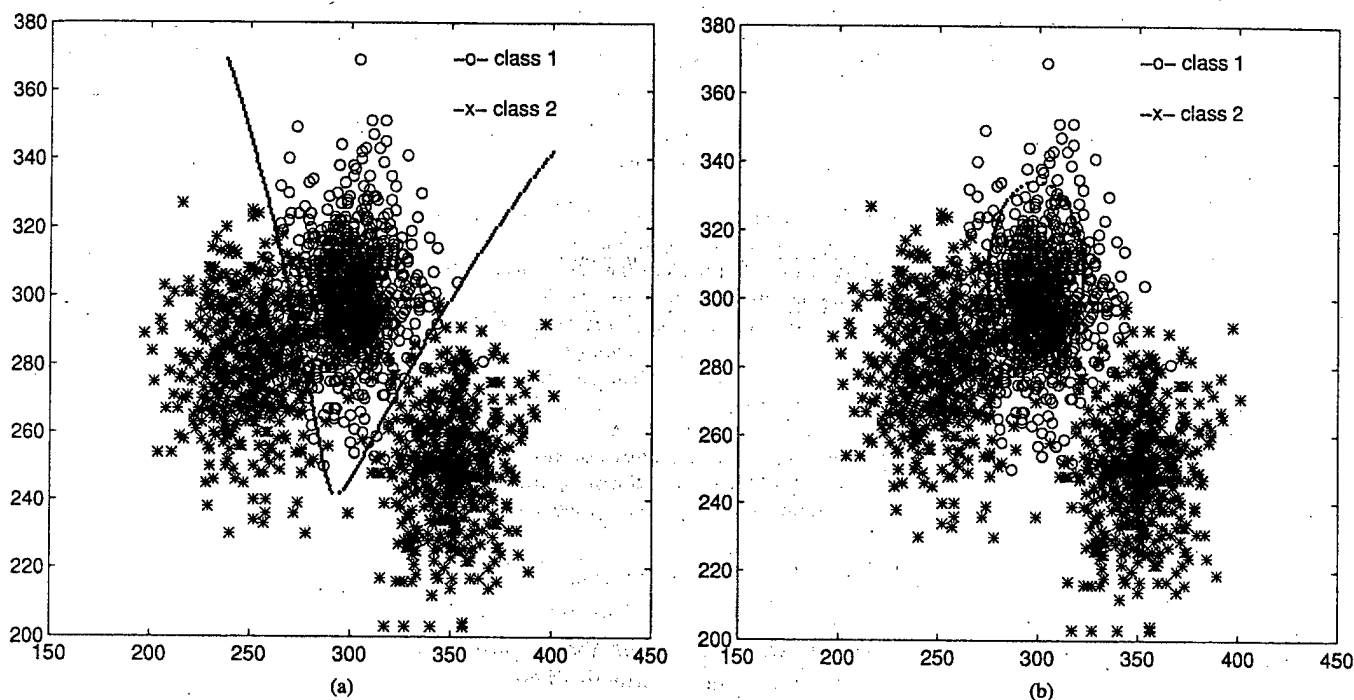


Fig. 6. The classification examples with a two-dimensional (2-D) simulated database (Database B, see Table II). (a) Class 2 contains two local clusters. (b) Class 2 contains one local cluster.

learning. Otherwise pattern \vec{x}_i belongs to the unknown class. We observed consistent and significant improvement in classification results compared with the pure Bayesian decision. Using the ORL (Olivetti Research Laboratory, Cambridge, U.K.) standard database (Database C, see Table II), our experience has shown an increase of correct detection rate from 70% to 90% [14].

In the third experiment, we use the proposed classifier to distinguish true masses from false masses based on the features extracted from the suspected regions. The objective is to reduce the number of suspicious regions and identify the true masses. 150 mammograms, each of them contains at least one mass case of varying size and location, were selected in our study. The areas of suspicious masses were identified following the proposed procedure with biopsy proven results. Fifty mammograms with biopsy proven masses were selected from the 150 mammograms for training (Database D, see Table II). The mammogram set used for testing contained 46 single-view mammograms: 23 normal cases and 23 with biopsy proven masses (Database E, see Table II) which were also selected from the 150 mammograms. All mammograms were digitized with an image resolution of $100 \mu\text{m} \times 100 \mu\text{m}/\text{pixel}$ by the laser film digitizer (Model: Lumiscan 150). The image sizes are $1792 \times 2560 \times 12$ bpp. For this study, we shrunk the digital mammograms with the resolution of $400 \mu\text{m}$ by averaging 4×4 pixels into one pixel. According to radiologists, the size of the small masses is 3–15 mm. The middle size of masses is 15–30 mm. The large size of masses is 30–50 mm, which are rare in mammograms. A 3-mm object in an original mammogram occupies 30 pixels in a digitized image with a $100\text{-}\mu\text{m}$ resolution. After reducing the image size by four times, the object will occupy the range of about seven to eight pixels. The object with the size of seven pixels is expected to be detectable by any computer algorithm.

Therefore, the shrinking step is applicable for mass cases and can save computation time.

After the segmentation, the area index feature was first used to eliminate the nonmass regions. In our study, we set $A_1 = 7 \times 7$ pixels and $A_2 = 75 \times 75$ pixels as the thresholds. A_1 corresponds to the smallest size of masses (3 mm), and an object with a area of 75×75 pixels corresponds to 30 mm in the original mammogram. This indicates that the scheme can detect all masses with sizes up to 30 mm. Masses larger than 30 mm are rare cases in the clinical setting. When the segmented region satisfied the condition $A_1 \leq A \leq A_2$, the region was considered to be suspicious for mass. For the purpose of representative demonstration, we have selected a 3-D feature space consisting of compactness I, compactness II, and difference entropy. According to our investigation, these three features have the better separation (discrimination) between the true and false mass classes. It should be noticed that the feature vector can easily extend to higher dimensionality. A training feature vector set was constructed from 50 true mass ROIs and 50 false mass ROIs (Database D, see Table II). The training set was used to train two modular probabilistic decision-based neural networks separately. In addition to the decision boundaries recommended by the computer algorithms, a visual explanation interface has also been integrated with 3-D to 2-D hierarchical projections. Fig. 7(a) shows the database map projection with compactness definition I and difference entropy. Fig. 7(b) shows the database map projection with compactness definition II and difference entropy. Our experience has suggested that the recognition rate with compactness I are more reliable than that with compactness II. In order to have more accurate texture information, the computation of the second-order joint probability matrix $p_{d, \theta}(i, j)$ is only based on the segmented region of the original mammogram. For the shrunk mammograms, we found that

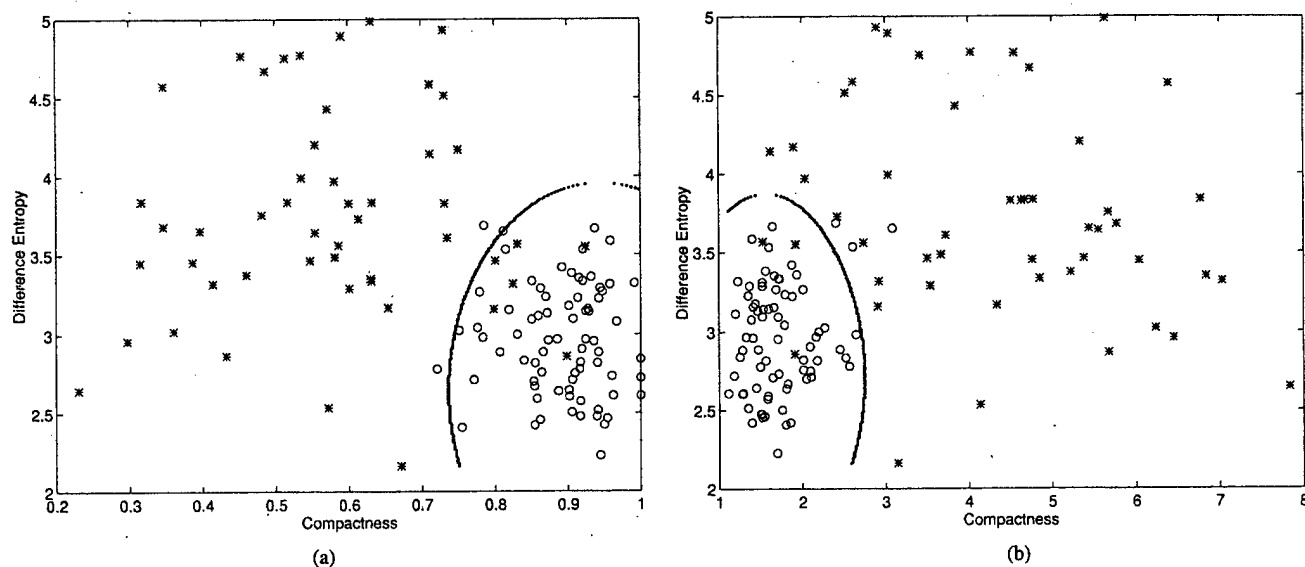


Fig. 7. The data mapping results (Database D, see Table II). -o- denotes true mass cases; -* denotes false mass cases. (a) The mapping using compactness I. (b) The mapping using compactness II.

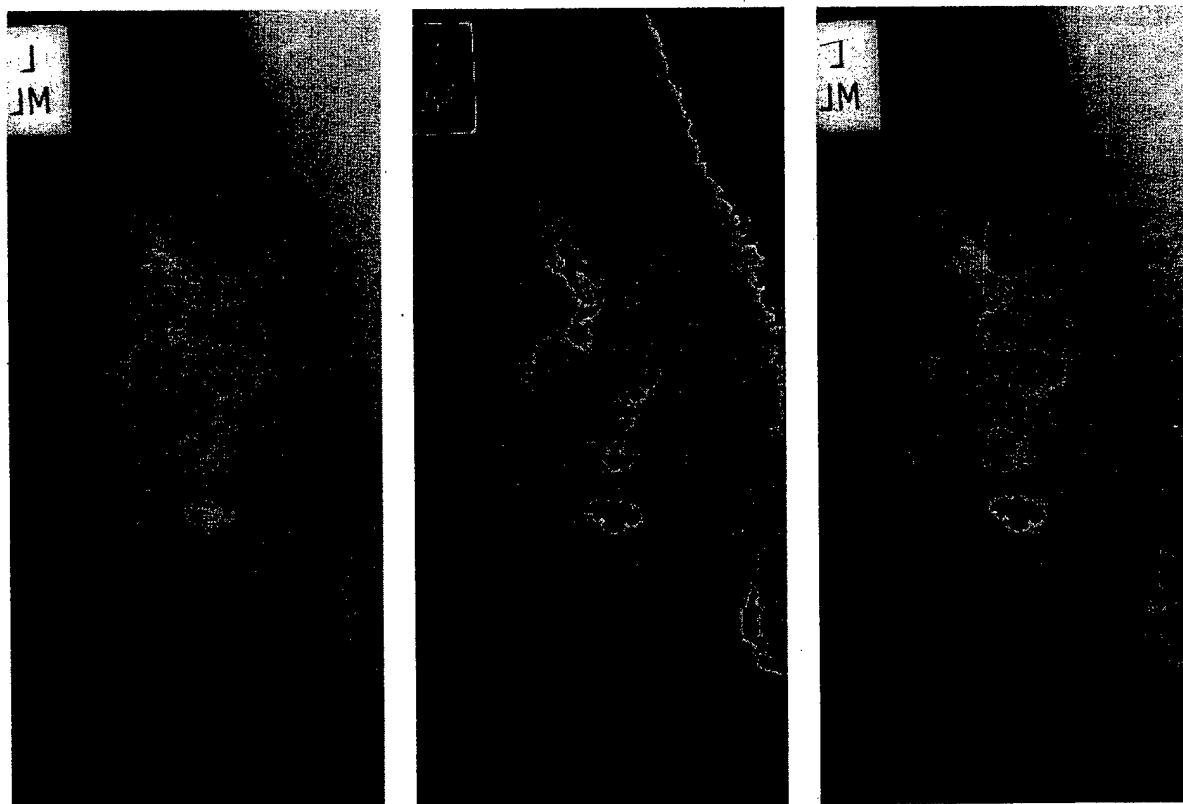


Fig. 8. One example of the mass detection using the proposed approach (Database E, see Table II).

the difference entropy had better discrimination with $d = 1$. The difference entropy used in this study was the average of values at $\theta = 0^\circ, 45^\circ, 90^\circ$, and 135° .

We have conducted a preliminary study to evaluate the performance of the algorithms in real case detection, in which 6–15 suspected masses/mammogram were detected and required further clinical decision making. We found that the proposed classifier can reduce the number of suspicious masses with a sensitivity of 84% at 1.6 false positive findings/mammogram based on the testing data set containing 46 mammograms (23 of them

have biopsy proven masses) (Database E, see Table II). Fig. 8 shows a representative mass detection result on one mammogram with a stellate mass. After the enhancement, ten regions with brightest intensity were segmented. Using the area criterion, too large and too small regions were eliminated first and the rest regions were submitted to the PMNN for further evaluation. The results indicated that the stellate mass lesion was correctly detected.

For further evaluation, receiver operating characteristic (ROC) method may be employed. However, we do not feel

ROC analysis will provide really a better evaluation but an alternative method to this case. First, most ROC analysis reported by others were based on different database thus are not comparable since ROC results are highly data-dependent. Second, ROC analysis only indicate an "overall" performance with limitations at least in twofold: it is for multithreshold thus the corresponding system may not be optimal to a particular application where only one threshold is needed; and it cannot provide a mathematically traceable feedback to improve the performance of the system or the one component in the system. Third, currently used FROC analysis package imposes several assumptions on the distributions of the cases which are invalid in most applications and particularly untrue in our situation. For example, our assumptions about the data distributions is SFNM that is clearly different from the restricted conditions imposed by the application of existing FROC analysis algorithm. In our approach, a quantitative mapping of the knowledge database is performed with hierarchical SFMD modeling and should be perfectly (at least in the theoretical sense) carried out by the corresponding PMNN classifier. In other words, optimal decision making should have already been achieved according to the Bayesian rule. It is reasonable to acknowledge that in order to compare the overall performance with the other systems, an ROC study may be further conducted. We are currently working on developing a new generation of FROC analysis package with a caution to remove the forementioned problems.

Another important consideration with the present approach is the measure of quality in visual explanation [29]. This is not a glamorous area, but progress in this area is eminently critical to the future success of visual exploration [28]. What is the correct matrix for a direct projection of a particular multimodal data set? How effective was a particular visualization tool? Did the user come to the correct conclusion? It may be agreeable that the benchmark criteria in visual exploration are very different and difficult [28]. As shared by Bishop and Tipping [27], we believe that in data visualization there is no objective measure of quality, and so it is difficult to quantify the merit of a particular data visualization technique, and the effectiveness of such a techniques is often highly data-dependent. The possible alternative is to perform a rigorous psychological evaluation using simple and controlled environment, or to invite domain experts to direct evaluate the efficacy of the algorithm for a specified task. For example, we can compare the domain expert's performances with and without the system aid. In that case, the ROC method may be used to evaluate the performance of our algorithm when used by the radiologists. While the optimality of these new techniques is often highly data-dependent, we would expect the hierarchical visualization model to be a very effective tool for the data visualization and exploration in many applications.

In summary, we employed a mathematical feature extraction procedure to construct the featured knowledge database from all the suspicious mass sites localized by the enhanced segmentation. The optimal mapping of the data points was then obtained by learning the generalized normal mixtures and decision boundaries. A visual explanation of the decision making was further invented as a decision support, based on an interactive

visualization hierarchy through the probabilistic principal component projections of the knowledge database and the localized optimal displays of the retrieved raw data. A prototype system was developed and pilot tested to demonstrate the applicability of this framework to mammographic mass detection.

ACKNOWLEDGMENT

The authors would like to thank R. F. Wagner of the Food and Drug Administration and S.-Y. Kung of the Princeton University for their valuable scientific input.

REFERENCES

- [1] R. Zwiggelaar, T. C. Parr, J. E. Schumm, I. W. Hutt, C. J. Taylor, S. M. Astley, and C. R. M. Boggis, "Model-based detection of spiculated lesions in mammograms," *Med. Image Anal.*, vol. 3, no. 1, pp. 39-62, 1999.
- [2] N. Karssemeijer and G. M. te Brake, "Detection of stellate distortions in mammogram," *IEEE Trans. Med. Imag.*, vol. 15, pp. 611-619, Oct. 1996.
- [3] L. Miller and N. Ramsey, "The detection of malignant masses by non-linear multiscale analysis," *Excerpta Medica*, vol. 1119, pp. 335-340, 1996.
- [4] N. Petrick, H. P. Chan, B. Sahiner, M. A. Helvie, M. M. Goodsitt, and D. D. Adler, "Computer-aided breast mass detection: False positive reduction using breast tissue composition," *Excerpta Medica*, vol. 1119, pp. 373-378, 1996.
- [5] W. K. Zouras, M. L. Giger, P. Lu, D. E. Wolverton, C. J. Vyborny, and K. Doi, "Investigation of a temporal subtraction scheme for computerized detection of breast masses in mammograms," *Excerpta Medica*, vol. 1119, pp. 411-415, 1996.
- [6] M. Zhang, M. L. Giger, C. J. Vyborny, and K. Doi, "Mammographic texture analysis for the detection of spiculated lesions," *Excerpta Medica*, vol. 1119, pp. 347-351, 1996.
- [7] W. P. Kegelmeyer Jr., J. M. Pruneda, P. D. Bourland, A. Hillis, M. W. Riggs, and M. L. Nipper, "Computer-aided mammographic screening for spiculated lesions," *Radiology*, vol. 191, pp. 331-337, 1994.
- [8] R. N. Strickland, "Tumor detection in nonstationary backgrounds," *IEEE Trans. Med. Imag.*, vol. 13, pp. 491-499, June 1994.
- [9] H. P. Chan, D. Wei, M. A. Helvie, B. Sahiner, D. D. Alder, M. M. Goodsitt, and N. Petrick, "Computer-aided classification of mammographic masses and normal tissue: Linear discriminant analysis in texture feature space," *Phys. Med. Biol.*, vol. 40, pp. 857-876, 1995.
- [10] M. L. Giger, C. J. Vyborny, and R. A. Schmidt, "Computerized characterization of mammographic masses: Analysis of spiculation," *Cancer Lett.*, vol. 77, pp. 201-211, 1994.
- [11] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [12] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610-621, Nov. 1973.
- [13] R. Schalkoff, *Pattern Recognition: Statistical, Structural, and Neural Approaches*. New York: Wiley, 1992.
- [14] Y. Wang, S. H. Lin, H. Li, and S. Y. Kung, "Data mapping by probabilistic modular networks and information theoretic criteria," *IEEE Trans. Signal Processing*, vol. 46, pp. 3378-3397, Dec. 1998.
- [15] L. Perlovsky and M. McManus, "Maximum likelihood neural networks for sensor fusion and adaptive classification," *Neural Networks*, vol. 4, pp. 89-102, 1991.
- [16] H. Gish, "A probabilistic approach to the understanding and training of neural network classifiers," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, 1990, pp. 1361-1364.
- [17] D. M. Titterton, A. F. M. Smith, and U. E. Markov, *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley, 1985.
- [18] C. E. Priebe, "Adaptive mixtures," *J. Amer. Stat. Assoc.*, vol. 89, no. 427, pp. 910-912, 1994.
- [19] S. Haykin, *Neural Networks: A Comprehensive Foundation*. New York: MacMillan College, 1994.
- [20] M. I. Jordan and R. A. Jacobs, "Hierarchical mixture of experts and the EM algorithm," *Neural Computation*, vol. 6, pp. 181-214, 1994.
- [21] S. Y. Kung and J. S. Taur, "Decision-based neural networks with signal/image classification applications," *IEEE Trans. Neural Networks*, vol. 1, pp. 170-181, Jan. 1995.

- [22] S. H. Lin, S. Y. Kung, and L. J. Lin, "Face recognition/detection by probabilistic decision-based neural network," *IEEE Trans. Neural Networks (Special issue on Artificial Neural Networks and Pattern Recognition)*, vol. 8, Jan. 1997.
- [23] Y. Wang, L. Luo, H. Li, and M. T. Freedman, "Hierarchical minimax entropy modeling and probabilistic principal component visualization for data explanation and exploration," presented at the SPIE Medical Imaging Conf., San Diego, CA, Feb. 20–26, 1999.
- [24] H. Li, Y. Wang, K. J. R. Liu, S.-C. B. Lo, and M. T. Freedman, "Computerized Radiographic Mass Detection—Part I: Lesion Site Selection by Morphological Enhancement and Contextual Segmentation," *IEEE Trans. Med. Imag.*, vol. 20, no. 4, pp. 289–301, Apr. 2001.
- [25] T. W. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [26] H. V. Poor, *An Introduction to Signal Detection and Estimation*. Berlin, Germany: Springer-Verlag, 1988.
- [27] C. M. Bishop and M. E. Tipping, "A hierarchical latent variable model for data visualization," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 281–293, Mar. 1998.
- [28] G. M. Nielson, "Challenges in visualization research," *IEEE Trans. Visual Comput. Graphics*, vol. 2, pp. 97–99, 1996.
- [29] E. R. Tufte, *Visual Explanation: Images and Quantities, Evidence and Narrative*. Cheshire, U.K.: Graphics, 1996.

Magnetic Resonance Image Analysis by Information Theoretic Criteria and Stochastic Site Models

Yue Wang, *Member, IEEE*, Tülay Adalı, *Member, IEEE*, Jianhua Xuan, and Zsolt Szabo

Abstract—Quantitative analysis of magnetic resonance (MR) images is a powerful tool for image-guided diagnosis, monitoring, and intervention. The major tasks involve tissue quantification and image segmentation where both the pixel and context images are considered. To extract clinically useful information from images that might be lacking in prior knowledge, we introduce an *unsupervised* tissue characterization algorithm that is both statistically principled and patient specific. The method uses adaptive standard finite normal mixture and inhomogeneous Markov random field models, whose parameters are estimated using expectation-maximization and relaxation labeling algorithms under information theoretic criteria. We demonstrate the successful applications of the approach with synthetic data sets and then with real MR brain images.

Index Terms—Finite normal mixture, image segmentation, information theoretic criteria, patient site model, tissue quantification.

I. INTRODUCTION

QUANTITATIVE analysis of magnetic resonance (MR) images refers to the problem of estimating tissue quantities and segmenting the anatomy into contiguous regions of interest. The problem has recently received much attention largely due to the improved fidelity and resolution of MR imaging systems, and the effective clinical utility of image analysis and understanding in diagnosis, monitoring, and intervention [1]–[4]. For example, pathological studies show that many neurological diseases are accompanied by subtle abnormal changes in brain-tissue quantities and volumes [2]. Due to the virtual impossibility for clinicians to quantitatively analyze these pathological changes associated with specific disease directly from MR images, considerable effort is required to develop accurate image analysis algorithms for identifying and quantifying these changes *in vivo* [5]–[10], [12].

The major tasks of MR image analysis involve tissue quantification and image segmentation. When a stochastic model-based

method is adopted, both pixel and context images should be considered. In this paper, we introduce an unsupervised image analysis procedure, which is both statistically principled and patient specific, a feature especially important in cases with limited or no prior knowledge [4], [10]. We assume that the MR images are single valued and the anatomy of the site may contain abnormalities. The method involves adaptive use of standard finite normal mixture (SFNM) and inhomogeneous Markov random field (MRF) models, whose parameters are estimated using the fast expectation-maximization (EM) and modified iterated conditional modes (MICM) algorithms under a selected information theoretic criterion [15]–[17]. The major difference of our study from previous research in the area [1]–[3], [5], [10] is as follows.

- 1) We introduce an adaptive SFNM to model the pixel images, where we use information theoretic criteria to determine the number of tissue types in a given image and use the boundary-associated pixels to fine tune the model parameter values. This allows incorporation of partial volume effect and tissue abnormality into a single stochastic model [3], [10].
- 2) We introduce an inhomogeneous MRF to model the context images, where we use the entropy rate distribution to determine probabilistic boundary sites in the image and the spatial discontinuity levels to assign the model parameter values. This permits an adaptive construction of the patient specific site model for improving both tissue quantification and image segmentation [23], [28].
- 3) We introduce post-global relative entropy (GRE) for the assessment of the final image-segmentation performance. Based on the observation that the parameter values of a particular tissue type in the quantified and segmented images should be very close, this criterion suggests an indirect, but objective approach for the difficult problem of performance evaluation in image segmentation [31].
- 4) We introduce a fast EM formulation, which is histogram based rather than conventionally raw data driven. Experiments show its very efficient performance.

In Section II, we discuss the stochastic modeling of both pixel and context images, and in Section III, we describe the methods for tissue quantification and image segmentation. This is further extended to demonstrate the accuracy and reproducibility of the algorithms in Section IV, where we first illustrate the operation of the algorithm along with the results using simulated data sets, and then apply the algorithm to real MR brain images. Finally, relationship of our method to other approaches and future work are discussed in Section V.

Manuscript received May 17, 2000. This work was supported in part by the Department of Defense under Grant DAMD17-98-8045 and by the National Institutes of Health under Grant 1R01AG14400.

Y. Wang is with the Department of Electrical Engineering and Computer Science, The Catholic University of America, Washington, DC 20064 USA and also with the Department of Radiology, Nuclear Medicine Division, Johns Hopkins Medical Institutions, Baltimore, MD 21205 USA (e-mail: wang@pluto.ee.cua.edu).

T. Adalı and J. Xuan are with the Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD 21250 USA.

Z. Szabo is with the Department of Radiology, Nuclear Medicine Division, Johns Hopkins Medical Institutions, Baltimore, MD 21205 USA.

Publisher Item Identifier S 1089-7771(01)02029-5.

II. IMAGE MODELING

In order to validate the use of a suitable stochastic model for MR image analysis with a specified objective, we have studied MR imaging statistics and observed several useful statistical properties of MR images [4]. These results are strongly supported by the analysis of actual MR image data [18]. In particular, based on the statistical properties of MR pixel images, where pixel image is defined as the observed gray level associated with the pixel, use of an SFNM distribution is justified to model the image histogram, and it is shown that the SFNM model converges to the true distribution when the pixel images are asymptotically independent [19]. Furthermore, by incorporating statistical properties of context images, where context image is defined as the membership of the pixel associated with different tissue types, we have proposed an inhomogeneous MRF formulation to impose local consistency constraints on context images in terms of a stochastic regularization scheme [20].

Assume that each pixel i can be decomposed into *pixel image* x_i and *context image* l_i with $i = 1, \dots, N$ [23]. By ignoring information regarding the spatial ordering of pixels, we can treat context images (i.e., pixel labels) as random variables and describe them using a multinomial distribution with unknown parameters π_k , where $\sum_{k=1}^K \pi_k = 1, 0 \leq \pi_k \leq 1$. Thus, the sufficient statistics are the pixel image statistics and the number of pixels for each component [15]. The marginal probability measure for any pixel image can be obtained by writing the joint probability density of x_i and l_i , and then summing the joint density over all possible outcomes of l_i , resulting in a sum of the following general form applicable to all pixels, according to the Bayes law

$$\begin{aligned} f(x) &= \sum_{k=1}^K \pi_k \frac{1}{\sqrt{2\pi} \sigma_k} \exp \left(-\frac{(x - \mu_k)^2}{2\sigma_k^2} \right) \\ &= \sum_{k=1}^K \pi_k g(x | \mu_k, \sigma_k^2) \end{aligned} \quad (1)$$

where μ_k and σ_k^2 are the mean and variance of the k th Gaussian kernel. We use K to denote the number of Gaussian components and $\mathbf{r} \in \mathcal{R}^{3K-1}$ to denote the model parameter vector that includes μ_k, σ_k^2 , and π_k for all K components (i.e., the tissue types). It is important to note that we consider the structural parameter K in (1) to be adjustable, not a fixed constant as assumed in most previous work [2], [3], [5]. We believe that this is an important consideration in unsupervised image analysis, better representing the cases encountered in the real clinical environment. For example, when prior knowledge regarding partial volume effect or tissue abnormality is not available for a given image, our proposed image model will adapt its structure to include all of the distinct components in the image [10], [13], [14].

In MR images, regions are piecewise smooth, thus, pixel label takes on discrete values and the labels of nearby pixels are strongly correlated. Using the equivalence between a Gibbs distribution and an MRF, it has been shown that Gibbs distribution provides a joint probability measure for context images \mathbf{l} in the following form [2], [10], [11], [23], [24]:

$$P(\mathbf{l}) = \frac{1}{Z_1} \exp(-U(\mathbf{l})) \quad (2)$$

where $U(\mathbf{l})$ is the energy function, and the normalizing constant Z_1 is the partition function. A neighborhood system can be established by specifying the clique function $V_c^{(i)}(\mathbf{l})$, where $U(\mathbf{l}) = \sum_{i=1}^N V_c^{(i)}(\mathbf{l})$. The most typical configuration of the MRF model is the pairwise interaction neighborhood system in which spatial neighbors occur only in pairs, i.e., the second-order model [23]. We define the neighborhood of pixel i , denoted by ∂i , by opening a 3×3 window with pixel i being the central pixel. The energy function can then be written as

$$U(\mathbf{l}) = \sum_{i=1}^N \sum_{j \in \partial i}^c [\theta_i I(l_i, l_j)] \quad (3)$$

where $\{\theta_i\}$ is the Markov parameter, c denotes clique with $c = 4$ for a second-order model, and $I(\cdot, \cdot)$ is the indicator function [4]

$$I(a, b) = \begin{cases} 1, & \text{if } a = b \\ 0, & \text{if } a \neq b. \end{cases} \quad (4)$$

By translating local context information into the energy function of the Gibbs measure, through clique structure or Markov parameter, an inhomogeneous MRF model can be established. Again, important to note that, in our formulation, the Markov parameter θ_i in (3) is considered to be shift variant, not a fixed constant, as in most previous work [2], [10]. We believe that the correlation among context images is primarily local and should be reflected by the Markov parameter values. For example, within one region, the context exhibits strong spatial continuity, while on region boundaries, the context reflects the natural discontinuity.

Simply applying the Bayes law, we can construct a unified framework to integrate the pixel image model with the context image model. It has been shown that requiring the conditional independence of the observed pixel images, given the context images, is sufficient to ensure that the posterior distribution is an MRF [21], [11], [22]–[24]. That is, a posterior conditional probability function for the context images \mathbf{l} , given the observed pixel images \mathbf{x} , has the form of a Gibbs distribution given by

$$P(\mathbf{l} | \mathbf{x}) = \frac{1}{Z_{1|\mathbf{x}}} \exp(-U(\mathbf{l} | \mathbf{x})) \quad (5)$$

where $Z_{1|\mathbf{x}}$ is a normalizing constant. Based on the pairwise interaction neighborhood system, the corresponding energy function is

$$\begin{aligned} U(\mathbf{l} | \mathbf{x}) &= \sum_{i=1}^N \left\{ \prod_{k=1}^K \left[\frac{1}{2} \ln(\sigma_k^2) + \frac{(x_i - \mu_k)^2}{2\sigma_k^2} \right]^{I(l_i, k)} \right. \\ &\quad \left. + \sum_{j \in \partial i}^c [\theta_i I(l_i, l_j)] \right\} \end{aligned} \quad (6)$$

where we have used (1) and (3).

Equation (5) refers to a hidden MRF, where the local property can be derived as

$$p(l_i | l_{\partial i}, \mathbf{x}) = \frac{1}{Z_i} \exp(-U(l_i | l_{\partial i}, \mathbf{x})) \quad (7)$$

where Z_i is a normalizing constant, $I_{\partial i}$ are local context images, and

$$U(l_i | I_{\partial i}, \mathbf{x}) = \prod_{k=1}^K \left[\frac{1}{2} \ln(\sigma_k^2) + \frac{(x_i - \mu_k)^2}{2\sigma_k^2} \right]^{I(l_i, k)} + \sum_{j \in \partial i} [\theta_j I(l_i, l_j)]. \quad (8)$$

The spatial statistical dependence among pixel images is one of the fundamental issues in the problem formulation for image analysis. In our approach, we tackle the problem as follows. For the purpose of tissue quantification, maximum likelihood (ML) estimation is used based on the SFNM model given in (1). In [19], we prove a convergence theorem showing that, when the pixel images are asymptotically independent, the parameter estimates based on (1) converge to their true values with probability one for sufficiently large N . For the purpose of image segmentation, a maximum a posterior probability (MAP) approach is employed to update \mathbf{l} according to (5). Since in the formulation we allow θ_i to be adjustable [21], the assignment of their values will incorporate the unified correlation among both the pixel and context images. Our tests with a wide class of simulated and real data have demonstrated the plausibility of the approach [32]. We include some of the examples in Section IV.

III. METHOD

A. Tissue Quantification

Tissue quantification addresses the combined estimation of regional parameters (π_k, μ_k, σ_k^2) and the detection of the structural parameter K in (1) given the pixel images \mathbf{x} . The two main approaches are based on classification-estimation and distance-minimization algorithms [2], [3], [6], [10]. In the classification-estimation approach, all pixels are first classified into different tissue types, and the model parameters are then estimated using the sample averages. In the distance-minimization approach, the mixture density is fitted to the histogram of pixel images by finding the optimal parameters. We use relative entropy (the Kullback–Leibler distance) for tissue quantification, a measure of the information theoretic distance between the histogram of the pixel images $f_{\mathbf{x}}$, and the estimated SFNM distribution $f(x)$, [25]

$$D(f_{\mathbf{x}} \| f) = \sum_{x \in \mathcal{X}} f_{\mathbf{x}}(x) \log \frac{f_{\mathbf{x}}(x)}{f(x)}. \quad (9)$$

We have shown that, when relative entropy is used as the distance measure, distance minimization is equivalent to the ML estimation of SFNM parameters under a pixel independency approximation, i.e., $\mathcal{L}(\mathbf{r}) = \exp(-N[H(f_{\mathbf{x}}) + D(f_{\mathbf{x}} \| f)])$, where H is the entropy calculator [19].

Thus, the ML estimate can be reached by minimizing $D(f_{\mathbf{x}} \| f)$, e.g.,

$$\frac{\partial D(f_{\mathbf{x}} \| f)}{\partial (\pi_k, \mu_k, \sigma_k^2)} = 0 \quad (10)$$

which motivates the consideration of a fast implementation of the EM algorithm based on the histogram $f_{\mathbf{x}}(x)$ rather than the

raw data \mathbf{x} , as formulated in [10] and [13]. We define the histogram of the data set as the number of times the value u_l occurs in the observation \mathbf{x}

$$f_{\mathbf{x}}(u_l) = \sum_{i=1}^N I(x_i, u_l) \quad (11)$$

for all $l = 1, \dots, L$ with $\sum_{l=1}^L f_{\mathbf{x}}(u_l) = N$. The new formulation is derived by augmenting the observed data histogram $f_{\mathbf{x}}(u_l)$ by splitting the value of $f_{\mathbf{x}}(u_l)$ at each level l into K_0 fractions of latent or unobserved data based on current estimates via posterior Bayes probability z_{lk} , as we now sketch the following steps:

E-Step

$$z_{lk}^{(n)} = \frac{\pi_k^{(n)} g(u_l | \mu_k^{(n)}, \sigma_k^{2(n)})}{f(u_l | \pi_k^{(n)}, \mu_k^{(n)}, \sigma_k^{2(n)})} \quad (12)$$

M-Step

$$\pi_k^{(n+1)} = \frac{1}{N} \sum_{l=1}^L z_{lk}^{(n)} f_{\mathbf{x}}(u_l) \quad (13)$$

$$\mu_k^{(n+1)} = \frac{\sum_{l=1}^L z_{lk}^{(n)} f_{\mathbf{x}}(u_l) u_l}{\sum_{l=1}^L z_{lk}^{(n)} f_{\mathbf{x}}(u_l)} \quad (14)$$

$$\sigma_k^{2(n+1)} = \frac{\sum_{l=1}^L z_{lk}^{(n)} f_{\mathbf{x}}(u_l) (u_l - \mu_k^{(n+1)})^2}{\sum_{l=1}^L z_{lk}^{(n)} f_{\mathbf{x}}(u_l)} \quad (15)$$

Although the result expressed by (12)–(15) may be intuitively obvious, i.e., by simply applying the expectation laws or reorganizing conventional formulation of the EM algorithm, it is also a formal result that follows from the philosophy of the EM algorithm presented in [15] with a new objective function $D(f_{\mathbf{x}} \| f)$. The algorithm first calculates the posterior Bayesian probabilities of the data through the observations and the current parameter estimates (*E*-step) and then updates parameter estimates using generalized mean ergodic theorems (*M*-step). The procedure cycles back and forth between these two steps until it reaches a stationary point of the likelihood, where n is the iteration index. In this application, we have experienced $L \ll N$, thus, the new algorithm will require $O(L)$ rather than $O(N)$ operations per iteration. For example, for most MR images, we have $L = 256 \sim 4096$ and $N = 65\,536 \sim 262\,144$.

For an unsupervised tissue quantification process, the number of tissue types should be detected as part of the parameter estimation procedure prior to image segmentation. This step is called model selection, which is particularly critical in clinical applications where the abnormality shown in the images for a particular patient may be arbitrarily complex. One approach to determine the optimal number K_0 is to use information theoretic criteria, such as the Akaike information criterion (AIC) [26], and the minimum description length (MDL) [27]. The major thrust of this approach has been the formulation of an adaptive

process in which a model fitting procedure is utilized to select a model from several competing candidates such that the selected model best fits the observed data.

For example, AIC will select the model that gives the minimum of

$$\text{AIC}(K_a) = -2 \log(\mathcal{L}(\hat{r}_{\text{ML}})) + 2K_a \quad (16)$$

where $\mathcal{L}(\hat{r}_{\text{ML}})$ is the likelihood of \hat{r}_{ML} , the ML parameter estimates, and K_a is the number of free adjustable parameters in the model. From a quite different point-of-view, MDL reformulates the problem explicitly as an information coding problem, in which the best model fit is measured such that high probabilities are assigned to the observed data, while at the same time, the model itself is not too complex to describe. The model is selected by minimizing the total description length defined by

$$\text{MDL}(K_a) = -\log(\mathcal{L}(\hat{r}_{\text{ML}})) + 0.5K_a \log N. \quad (17)$$

Note that, different from AIC, the second term in MDL takes into account the number of observations.

B. Image Segmentation

Image segmentation is a technique for partitioning the image into meaningful regions corresponding to different objects. It may be considered to be a clustering process where the pixels are classified into attributed tissue types according to their gray-level values and spatial correlation. More precisely, image segmentation addresses the realization of context images \mathbf{l} , given the observed pixel images \mathbf{x} [17]. Based on the inhomogeneous MRF model given by (5), there are several approaches for performing the pixel classification. For example, we can use ML classification to directly maximize the individual likelihood function of the pixel images, i.e., the first term in (5), by searching the optimum \mathbf{l} where the true pixel labels \mathbf{l}^* are considered to be functionally independent constants [17]. The major problem associated with this approach is that the classification error is high when the observed pixel images are noisy. However, it may well function as an initial solution since the classification error is spatially uniformly distributed. By considering both terms in (8), we derive the MICM algorithm to search for a MAP image segmentation. The structure of this relaxation labeling procedure is based on the following two basic considerations: 1) decomposition of the global computation scheme into a network performing simple local computations and 2) use of suitable local context regularities in resolving ambiguities [17].

Let ϵ denote the expected segmentation error, i.e., the posterior cost of misclassification given pixel images \mathbf{x} , based on the following equivalence:

$$\arg \left\{ \max_k P(l_i = k, \mathbf{l}_{S|i} | \mathbf{x}) \right\} = \arg \left\{ \max_k p(l_i = k | \mathbf{l}_{S|i}, \mathbf{x}) \right\} \quad (18)$$

where $\mathbf{l}_{S|i}$ denotes pixel labels of all pixels, except pixel i . References [22]–[24] show that choosing the labeling that minimizes ϵ is equivalent to maximizing the marginal posterior distribution such that label l_i is updated to satisfy

$$p(l_i^{(m+1)} | \mathbf{x}, \mathbf{l}_{S|i}) \geq p(l_i^{(m)} | \mathbf{x}, \mathbf{l}_{S|i}) \quad (19)$$

for all pixels. Furthermore, by imposing the Markovian constraint, we have the following relationship [22], [23]:

$$p(l_i | \mathbf{x}, \mathbf{l}_{S|i}) \propto p(x_i | l_i) p(l_i | \mathbf{l}_{\partial i}) \propto p(l_i = k | x_i, \mathbf{l}_{\partial i}). \quad (20)$$

That is,

$$\begin{aligned} \arg \left\{ \max_k p(l_i | \mathbf{x}, \mathbf{l}_{S|i}) \right\} &= \arg \left\{ \max_k p(x_i | l_i) p(l_i | \mathbf{l}_{\partial i}) \right\} \\ &= \arg \left\{ \max_k p(l_i | x_i, \mathbf{l}_{\partial i}) \right\}. \end{aligned} \quad (21)$$

Hence, based on the inhomogeneous hidden MRF formulation, MICM algorithm is constructed as a computationally feasible alternative to obtain an MAP solution. From (7), we can show that maximizing the conditional probability is equivalent to minimizing the energy function $U(l_i | \mathbf{l}_{\partial i}, \mathbf{x})$. That is, pixel i will be classified into the k th region if

$$l_i = \arg \left\{ \min_k U(l_i = k | \mathbf{l}_{\partial i}, \mathbf{x}) \right\}. \quad (22)$$

The MICM algorithm uses a specified number of iterations by randomly visiting all pixels. At each single cycle in the iterations, for the update of l_i , we have

$$P(\mathbf{l} | \mathbf{x}) = p(l_i | \mathbf{l}_{S|i}, \mathbf{x}) P(\mathbf{l}_{S|i} | \mathbf{x}) \quad (23)$$

which never decreases, hence, assuring eventual convergence [22], [23]. This approach demonstrates how a network of discrete units can be used to search an optimal solution in a problem where incorporation of context constraints is important.

To perform image segmentation based on the inhomogeneous MRF model requires assignment of the Markov parameter θ_i [23]. An important area in image analysis that recently started to be emphasized is the use of patient-specific site models to guide various medical image analysis tasks [28]. Hence, in the assignment of the Markov parameter, we first construct a patient-specific probabilistic site model directly from the interim context images (e.g., after ML classification), where the entropy rate distribution is used as a measure of the pixel label dependence, and then assign the Markov parameter values based on this model through a predefined lookup table. This approach has a clear physical interpretation since lower entropy rate means higher dependence and the probabilistic site model coincides with the boundary allocation [30].

For a stationary random process l_1, l_2, \dots, l_N , the entropy rate $H(\mathcal{L})$ is defined as the rate of entropy growth [25]. When dealing with a stationary Markov chain, the entropy rate can be calculated by

$$\begin{aligned} H(\mathcal{L}) &= \lim_{N \rightarrow \infty} H(l_N | l_{N-1}, \dots, l_1) \\ &= \lim_{N \rightarrow \infty} H(l_N | l_{N-1}) \\ &= H(l_2 | l_1). \end{aligned} \quad (24)$$

Assume that the inhomogeneous MRF is locally stationary, then the local entropy rate, i.e., the entropy rate distribution, is given by

$$H(\mathcal{L}_i) = \sum_{\mathbf{l}_{\partial i}} P(\mathbf{l}_{\partial i}) H(l_i | \mathbf{l}_{\partial i}). \quad (25)$$

Since the probability distribution of $l_{\partial i}$ is generally unavailable, we use first-order stochastic approximation to estimate the entropy rate [29], [30]

$$H(\mathcal{L}_i) \approx H(l_i | l_{\partial i}) = - \sum_{k=1}^K p(l_i = k | l_{\partial i}) \log p(l_i = k | l_{\partial i}). \quad (26)$$

By opening a 3×3 neighborhood window, the conditional probability is given by

$$p(l_i = k | l_{\partial i}) = \sum_{j \in \partial i} \frac{I(l_j, k)}{8}. \quad (27)$$

It can be shown that when $l_{\partial i}$ represents a uniform block, the corresponding entropy rate is zero. Finally, we use the following lookup table to assign the Markov parameter values

$$\theta_i = \frac{\alpha}{H(\mathcal{L}_i) + \eta} \quad (28)$$

where α is the scale factor and η is the shifting offset, and these coefficients are empirically determined based on our experience. We refer to the plot of θ_i as the patient-specific probabilistic site model or the site model.

Finally, we introduce, in addition to the iterated update of context images l using (22) and subsequent modification of the Markov parameter θ according to lookup table (28), the fine tuning of the tissue conditional likelihood densities through a "classification-based learning" [19]. The classification-based learning uses the *misclassified* pixels to adjust the tissue density functions, which are previously estimated using the EM algorithm, so that the minimum classification error can be achieved. Integrated into the MICM algorithm, the reinforced and antireinforced learning rules are used to update the tissue quantities μ_k and σ_k^2 in the first term of (6) at each iteration as follows:

Reinforced learning

$$\begin{aligned} \mu_{l_i^{(n+1)}}^{(n+1)} &= \mu_{l_i^{(n+1)}}^{(n)} + \kappa \left(x_i - \mu_{l_i^{(n+1)}}^{(n)} \right) \\ \sigma_{l_i^{(n+1)}}^{2(n+1)} &= \sigma_{l_i^{(n+1)}}^{2(n)} + \kappa \left[\left(x_i - \mu_{l_i^{(n+1)}}^{(n)} \right)^2 - \sigma_{l_i^{(n+1)}}^{2(n)} \right]. \end{aligned}$$

Anti-reinforced learning

$$\begin{aligned} \mu_{l_i^{(n)}}^{(n+1)} &= \mu_{l_i^{(n)}}^{(n)} - \kappa \left(x_i - \mu_{l_i^{(n)}}^{(n)} \right) \\ \sigma_{l_i^{(n)}}^{2(n+1)} &= \sigma_{l_i^{(n)}}^{2(n)} - \kappa \left[\left(x_i - \mu_{l_i^{(n)}}^{(n)} \right)^2 - \sigma_{l_i^{(n)}}^{2(n)} \right] \end{aligned} \quad (29)$$

where the misclassified pixels are those with $l_i^{(n+1)} \neq l_i^{(n)}$ and κ is the learning rate constant, typically chosen small ($\kappa \ll 1$). It is important to notice that, if the pixels are reclassified into $l_i^{(n+1)}$ th tissue type, reinforced learning will be applied to pull the corresponding kernel closer to the host region, while antireinforced learning will be used to push the kernel of $l_i^{(n)}$ th tissue type away from the problematic region [19]. The misclassified pixels associated with the spatial discontinuities are used to fine tune the discriminant thresholds in the decision domain. We refer to those misclassified pixels as the boundary-associated

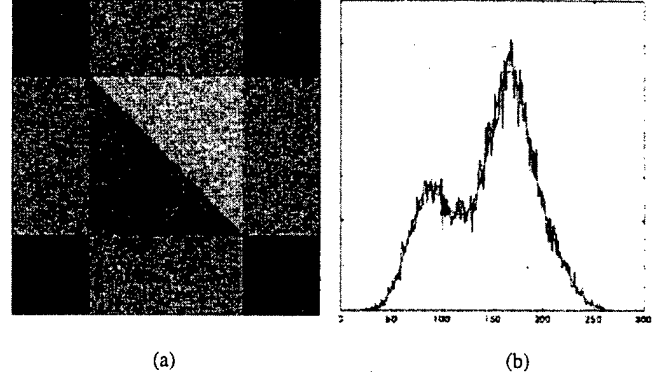


Fig. 1. Simulated image with: (a) four Gaussian components and the (b) corresponding histogram of the pixel images (right).

pixels since significant partial volume effect most often occurs along the boundaries [3].

IV. ALGORITHM AND EXPERIMENT

Thus far, we have described the theory behind the stochastic site model supported image analysis scheme with tissue quantification and segmentation. We now complete the description of our algorithm by considering the major steps in the implementation and illustrating its performance. Although the tissue quantification and image segmentation may be simultaneously performed [3], [5], [6], a more interesting effort, with greater accuracy, is to achieve these two objectives separately [4], [10], [23]. Guided by the two information theoretic criteria, our algorithm progressively proceeds by fitting a SFNM to the histogram of pixel images with model order selection, and then constructs a consistent relaxation labeling of the context images. A summary of the major steps is given as follows.

- Step 1) For each value of K , perform ML tissue quantification by the EM algorithm (12)–(15) with 30–50 iterations].
- Step 2) Scan the values of $K = K_{\min}, \dots, K_{\max}$, by AIC (16) and MDL (17) to determine the suitable number of tissue types K_0 .
- Step 3) Select the result of tissue quantification corresponding to the value of K_0 determined in step 2.
- Step 4) Initialize image segmentation by ML classification.
- Step 5) Construct the probabilistic site model or site model according to the entropy rate distribution (26), (27) and the lookup table (28).
- Step 6) Finalize image segmentation by MICM algorithm by implementing relaxation labeling (22) and local parameter tuning (29) with 5–10 iterations.

The performance of the algorithm is evaluated in terms of the GRE value for tissue quantification, visual judgement, and post-GRE value for image segmentation, as well as computational complexity and robustness [4], [22], [31].

We first illustrate the application of our algorithm to a simple synthetic data set generated from a mixture of four Gaussians in a two-dimensional space, shown in Fig. 1(a). Each component represents one tissue type. The value for each of the components is set to a constant value and normally distributed noise is added to the simulation phantom resulting in 6-dB SNR. Fig. 1(b)

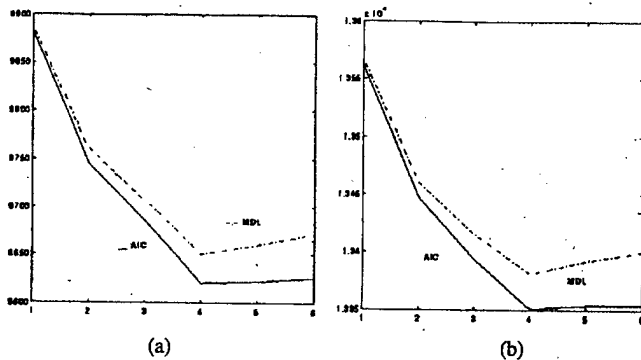


Fig. 2. Model selection with simulated image using: (a) AIC and (b) MDL (number in axis corresponds to the number of image regions).

TABLE I
TRUE AND ESTIMATED PARAMETER VALUES FOR THE SIMULATED
IMAGE IN FIG. 1

k	$\pi/\hat{\pi}$	$\mu/\hat{\mu}$	$\sigma^2/\hat{\sigma}^2$
1	0.25/0.24	86/84	400/354
2	0.125/0.13	126/121	400/365
3	0.5/0.48	166/164	400/373
4	0.125/0.15	206/201	400/463

shows the corresponding histogram of the pixel images (noisy curve). It is clear that these four simulated tissue types are highly overlapping. We then apply the two information theoretic criteria (AIC and MDL) to detect the number of the components. The curves of the AIC and MDL, as functions of the number of tissue types K , are plotted in Fig. 2(a) and (b), respectively, where the plots correspond to two different noise levels. The minima of these curves indicate the correct number of image components. It is clear that the number of tissue components suggested by both criteria is correct. The results of the distribution learning using the EM algorithm are shown in Fig. 1(b) (smooth curve). With the correct K_0 , the GRE between the estimated SFNM and the image histogram is about 0.008 nats. The numerical results are given in Table I.

Having determined the parameters of the SFNM model, we conduct image segmentation where we compare the performance of three different algorithms. The ML pixel classification is applied to initialize image segmentation, where the pixel is classified into the k th tissue component based on the ML, i.e.,

$$l_i = \arg \left\{ \min_k \prod_{k=1}^K \left[\frac{1}{2} \ln(\sigma_k^2) + \frac{(x_i - \mu_k)^2}{2\sigma_k^2} \right]^{I(l_i, k)} \right\}.$$

The result is shown in Fig. 3(a). The corresponding classification error is about 30% and is spatially and uniformly distributed over the whole image. We then apply the ICM algorithm to update the pixel labeling with two settings, i.e., $\theta = 0.4$ and $\theta = 6$. The results are shown in Fig. 3(b) and (c), respectively. It can be concluded that since the ICM algorithm uses a homogeneous MRF configuration, a smaller θ will lead to a noisy image segmentation with higher classification error, and a larger θ will lead to a smooth context recover with missing details. For the given simulated image, we determined $\theta = 1.5$ as the best choice, and with this setting, the corresponding classification error is down to 0.7508%.

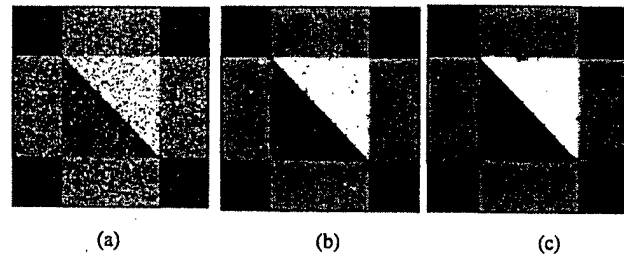


Fig. 3. Results of image segmentation using: (a) the ML pixel classification, (b) ICM algorithm after the third iteration with $\theta = 0.4$, and (c) the ICM algorithm after the third iterations with $\theta = 6$.

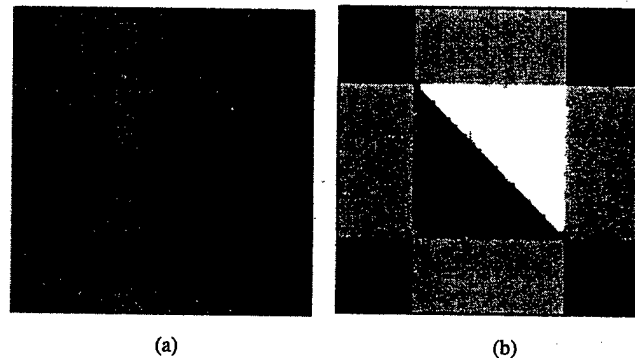


Fig. 4. Contextual image segmentation using a probabilistic site model [(a) highlight the edges of regions] (b) supported approach.

Finally, we use the MICM algorithm together with the Markov parameter assignment procedure to improve image segmentation. Based on (26)–(28), our intensive numerical experiments suggest that the assignment of $\theta = 0.4$ along the boundaries and $\theta = 1.5$ – 6.0 within the regions, results in good performance in a typical tone image [22], [23]. Using this empirical knowledge as a guideline, we select $\alpha = 0.5$ and $\eta = 0.1$ to construct the lookup table. Fig. 4(a) gives the distribution of the Markov parameter values for the simulated image, clearly showing the structure of the boundaries. The result using the MICM algorithm is shown in Fig. 4(b), which provides an almost “perfect” true context in which the image details are nicely preserved while most noise effects have been removed. The final classification error is further reduced to 0.3113%.

As an example of a more complex problem, we consider a set of real MR brain images used for identification of different tissue space domains [three-dimensional (3-D) spoiled gradient recalled acquisition (SPGR)]. A slice from such a data set is shown in Fig. 5(a). This is a T2-weighted image parallel to the ac-PC line. The data are acquired with a GE Sigma 1.5 Tesla system. The imaging parameters are TR 35, TE 5, flip angle 45° , 1.5-mm effective slice thickness, zero gap, and 124 slices with in-plane 192×256 matrix, and 24-cm field of view, and one number of excitation (nex). Since the skull, scalp, and fat in the original brain images do not contribute to the brain tissue, we edit the MR images to exclude nonbrain structures prior to tissue quantification and segmentation. A close look of the whole 3-D data set has indicated that the histogram has a considerably different characteristics from slice to slice and the tissue types are all highly overlapping [4]. Our experiment is to assess the feasibility of the algorithm on model selection and probabilistic site model construction with real MR images. After the initial image

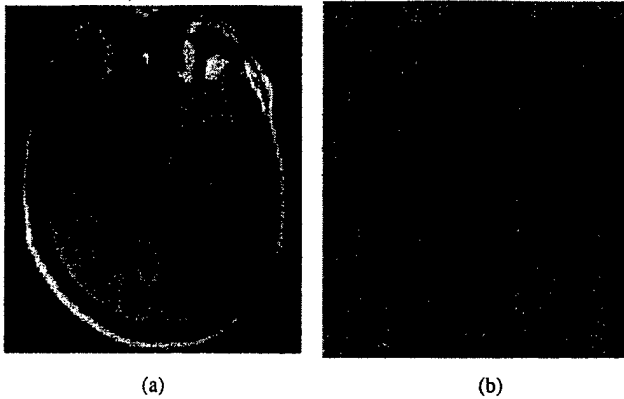


Fig. 5. (a) Real MR brain image and (b) the corresponding probabilistic site model, where the key feature is the natural discontinuities in the image.

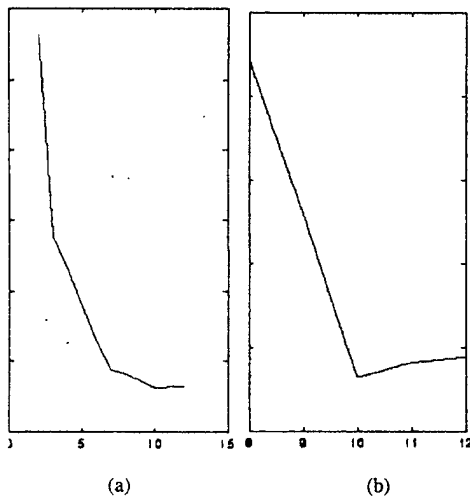


Fig. 6. MDL curve of the MR image in Fig. 5. (a) Global profile. (b) Local detail, where the minimum of the MDL value corresponds to $K_0 = 10$.

segmentation by the ML pixel classification, the corresponding site model is given in Fig. 5(b). Applying both AIC and MDL to this image, we obtain $K_0 = 10$. The corresponding MDL curve is shown in Fig. 6(a) with a close-up around the minimum, as shown in Fig. 6(b). Our experience has shown that the overall performance of these two information theoretic criteria is fairly consistent with real MR brain images.

As a final example, we consider tissue quantification and image segmentation of a normal T1-weighted MR brain image, as shown in Fig. 7(a). As discussed in the literature [2], [3], the brain tissue is generally composed of three principal types, i.e., white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), and their pairwise combinations. Santiago and Gage [3] propose a six-tissue model to represent the primary tissue types and the partial volume effect, defined as CSF-white (CW), CSF-gray (CG), and gray-white (GW). In our implementation, we also consider the triple mixture tissue, defined as CSF-white-gray (CWG). More importantly, since the MR images clearly show distinctive intensities at local brain areas, the functional areas within a tissue type need to be considered. In particular, the caudate nucleus and putamen are two important local brain functional areas. We calculate $AIC(K)$ and $MDL(K)$ for $K = 2, \dots, 10$. The results with these two criteria are given in Table II, which suggest that this brain image contains

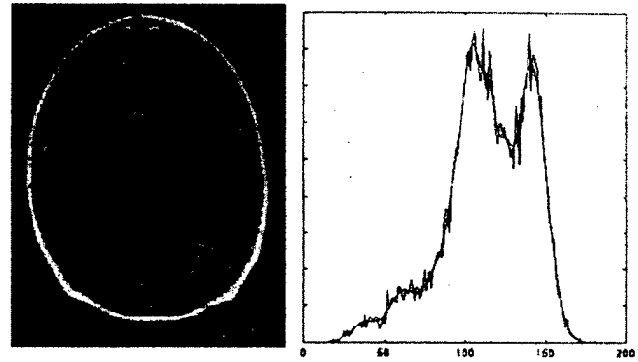


Fig. 7. (a) T1-weighted "normal" MR brain image and (b) the corresponding image histogram.

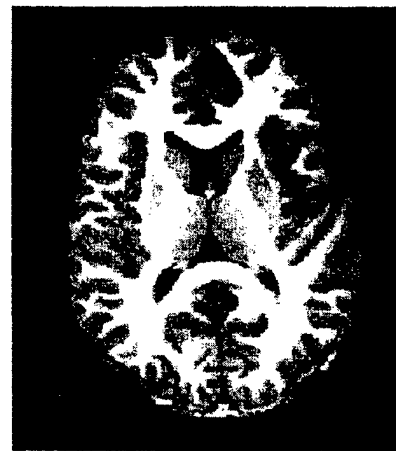


Fig. 8. Result of image segmentation using the MICM algorithm and local parameter fine tuning.

eight tissue types. The histogram of the pixel images is shown in Fig. 7(b) (noisy curve). With $K_0 = 8$, the result of tissue quantification using our algorithm is given by Fig. 7(b) (smooth curve) with a GRE value of 0.02–0.04 nats. The actual parameter values corresponding to the estimation agree with those of a physician's qualitative analysis results [4]. The MICM-based tissue segmentation together with the local parameter tuning (29) is performed and the pixel label updates are terminated after 5–10 iterations since further iterations produced almost identical results. The segmentation result are given in Fig. 8. Although the segmentation result contains some small isolated spots (less than four-pixel size), the proposed approach is quite encouraging. It is seen that the boundaries of WM, GM, and CSF are delineated very well and successfully. The regions with different gray levels are satisfactorily segmented, especially the major brain tissues, which are clearly identified and represent the following eight brain tissue types:

- 1) CSF;
- 2) CG;
- 3) CGW;
- 4) GW;
- 5) GM;
- 6) putamen area;
- 7) caudate area;
- 8) WM.

TABLE II
MODEL SELECTION USING AIC AND MDL VALUES

K	2	3	4	5	6	7	8	9	10
AIC	295944	291667	288845	288397	288380	288184	285685	287674	286317
MDL	295967	291703	288895	288461	288458	288275	285790	287792	286449

These segmented tissue types also agree with the results of radiologists' evaluation.

We have tested our algorithm with a total of 124 MR brain images. As illustrated here by both simulated and real examples, we have observed that the overall approach can be very effective at quantifying and segmenting different tissue types. It is important to emphasize that in unsupervised image analysis with real data sets, an objective evaluation of different techniques is a particularly difficult task [6], [10], [31]. It is difficult to quantify the merit of a particular algorithm, and dependability of evaluations by simple mathematical measures is questionable [4]. Therefore, most of the time, the quality of tissue quantification and image segmentation depends heavily on the subjective and qualitative judgments. In this paper, besides the evaluation performed by radiologists, we use the GRE value to measure the quality of tissue quantification. We believe that this criterion is quite reasonable since the ML estimation is unbiased under the SFNM modeling and image histogram, as the reference is the only subjective quantity directly available from the image data [15], [29]. Furthermore, for the assessment of image segmentation, we use both the post-GRE value and the visual inspection, where the post-GRE value is generated by the sample averages based on 1

$$\begin{cases} \hat{\pi}_k = \frac{1}{N} \sum_{i=1}^N I(l_i, k) \\ \hat{\mu}_k = \frac{1}{N\pi_k} \sum_{i=1}^N I(l_i, k) x_i \\ \hat{\sigma}_k^2 = \frac{1}{N\pi_k} \sum_{i=1}^N I(l_i, k) (x_i - \mu_k)^2 \end{cases} \quad (30)$$

and is given by

$$D(f_z \| f_1) = \sum_{x \in \mathcal{X}} f_z(x) \log \frac{f_z(x)}{f_1(x)} \quad (31)$$

as an indirect, but objective criterion, where $f_1(x)$ denotes the SFNM distribution corresponding to the segmented image. Considering the probabilistic pixel label z and the realized pixel label l , the pre-segmentation SFNM f_z (e.g., $f(x)$ estimated via a soft clustering procedure with EM) and post-segmentation SFNM f_1 are related through the image segmentation obtained by the MICM algorithm. The post-GRE segmentation criterion presented here states that the relative entropy between the ML estimate of the SFNM (via soft clustering) and the SFNM obtained from the segmented image (via hard clustering) is minimal if the image components are correctly segmented. The parameter values of a particular tissue type in the estimated histogram are most likely to be equal to the parameter values of the corresponding tissue type in the segmented region if the

TABLE III
COMPARISON OF SEGMENTATION ERROR RESULTING FROM DIFFERENT CLUSTERING METHOD

Method	ML-EM (soft)	ICM (hard-homogeneous)	MICM (hard-inhomogeneous)
GRE value (nats)	0.0067	0.4406	0.1578

pixel images are properly classified. This correspondence is lost in the case of misclassification. Our experiment with the MR brain image has led to a result of $D(f_x \| f_1) = 0.1578$ nats and $D(f_x \| f_z) = 0.0067$ nats, which coincides with what is suggested by the information theoretic Pythagorean theorem [25, Th. 12.6.1], $D(f_x \| f_1) \geq D(f_x \| f_z) + D(f_z \| f_1)$. This result implies that the classification error in image segmentation contributes to the additional quantification error $D(f_z \| f_1)$. Table III presents the summary of the results.

V. SUMMARY

In this paper, we have presented a complete procedure for quantifying and segmenting major brain tissue types from MR images. This method, as illustrated by well-planned simulation and pilot applications in brain-tissue analysis, is capable of revealing abnormalities within data (e.g., allowing K to be adjustable) and can be applied to clinical problems such as those encountered in tissue segmentation and quantitative diagnosis. It is important to emphasize that, although the experiments here are presented with two-dimensional images, their application to 3-D data sets is straightforward and has been implemented recently [32].

The main limitations of the current approach are that: 1) there is no systematic way to initialize the model parameter values and 2) the algorithm for both the tissue quantification and model selection is frequently trapped in a local optima. Thus, future effort should focus on a generic site model supported parameter initialization and a multiresolution or hierarchical scheme for both model selection and tissue quantification. Additional work includes theoretical investigation of the new formulation in model selection, partial volume effect by image blurring, and the objective criterion for image-segmentation evaluation.

REFERENCES

- [1] A. P. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer, "Morphometric analysis of white matter lesions in MR images: method and validation," *IEEE Trans. Med. Imag.*, vol. 13, pp. 716-724, Dec. 1994.
- [2] H. S. Choi, D. R. Haynor, and Y. Kim, "Partial volume tissue classification of multichannel magnetic resonance images—A mixture model," *IEEE Trans. Med. Imag.*, vol. 10, pp. 395-407, Sept. 1994.
- [3] P. Santago and H. D. Gage, "Quantification of MR brain images by mixture density and partial volume modeling," *IEEE Trans. Med. Imag.*, vol. 12, pp. 566-574, Sept. 1993.
- [4] Y. Wang, T. Adali, S.-Y. Kung, and Z. Szabo, "Quantification and segmentation of brain tissues from MR images: A probabilistic neural network approach," *IEEE Trans. Image Processing*, vol. 7, pp. 1165-1181, Aug. 1998.

- [5] H. E. Cline, W. E. Lorensen, R. Kikinis, and R. Jolesz, "Three-dimensional segmentation of MR images of the head using probability and connectivity," *J. Computer-Assisted Tomography*, vol. 14, pp. 1037-1045, 1990.
- [6] L. O. Hall, A. M. Bensaid, L. P. Clarke, R. P. Velthuizen, M. S. Silbiger, and J. C. Bezdek, "A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain," *IEEE Trans. Neural Networks*, vol. 3, pp. 672-682, Sept. 1992.
- [7] M. E. Brummer, R. M. Merserrau, R. L. Eisner, and R. J. Lewine, "Automatic detection of brain contours in MRI data sets," *IEEE Trans. Med. Imag.*, vol. 12, pp. 153-166, June 1993.
- [8] M. Joliot and B. M. Mazoyer, "Three-dimensional segmentation and interpolation of magnetic resonance brain images," *IEEE Trans. Med. Imag.*, vol. 12, pp. 269-277, June 1993.
- [9] Y. Wang, T. Adali, C.-M. Lau, and S.-Y. Kung, "Quantitative analysis of MR brain image sequences by adaptive self-organizing mixtures," *J. VLSI Signal Processing*, vol. 8, pp. 219-239, 1998.
- [10] Z. Liang, J. R. MacFall, and D. P. Harrington, "Parameter estimation and tissue segmentation from multispectral MR images," *IEEE Trans. Med. Imag.*, vol. 13, pp. 441-449, Sept. 1994.
- [11] S. Z. Li, *Markov Random Field Modeling in Computer Visions*. Berlin, Germany: Springer-Verlag, 1995.
- [12] J. Xuan, T. Adali, and Y. Wang, "Segmentation of magnetic resonance brain image: Integrating region growing and edge detection," in *Proc. 2nd IEEE Int. Image Processing Conf.*, vol. III, Washington, DC, 1995, pp. 544-547.
- [13] J. Zhang, J. W. Modestino, and D. A. Langan, "Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation," *IEEE Trans. Image Processing*, vol. 3, pp. 403-419, July 1994.
- [14] D. A. Langan, J. W. Modestino, and J. Zhang, "Cluster validation for unsupervised stochastic model-based image segmentation," *IEEE Trans. Image Processing*, vol. 7, pp. 180-195, Feb. 1998.
- [15] D. M. Titterton, A. F. M. Smith, and U. E. Markov, *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley, 1985.
- [16] R. A. Redner and N. M. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Rev.*, vol. 26, pp. 195-239, 1984.
- [17] R. A. Hummel and S. W. Zucker, "On the foundations of relaxation labeling processes," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-5, pp. 267-287, May 1983.
- [18] M. Fuderer, "The information content of MR images," *IEEE Trans. Med. Imag.*, vol. 7, pp. 368-380, Dec. 1988.
- [19] Y. Wang, S.-H. Lin, H. Li, and S.-Y. Kung, "Data mapping by probabilistic modular networks and information theoretic criteria," *IEEE Trans. Signal Processing*, vol. 46, pp. 3378-3397, Dec. 1998.
- [20] T. Adali, Y. Wang, and N. Gupta, "A blockwise relaxation labeling scheme and its application to edge detection in cardiac MR image sequences," in *International Journal of Imaging Science Technology*. New York: Wiley, 1998, vol. 9, pp. 340-350.
- [21] D. Geman, *Random Fields and Inverse Problems in Imaging*. Berlin, Germany: Springer-Verlag, 1990.
- [22] R. C. Dubes, A. K. Jain, S. G. Nadabar, and C. C. Chen, "MRF model-based algorithms for image segmentation," in *Proc. Pattern Recognition*, vol. 1, 1990, pp. 808-814.
- [23] R. C. Dubes and A. K. Jain, "Random field models in image analysis," *J. Appl. Stat.*, vol. 16, no. 2, pp. 131-164, 1989.
- [24] J. Besag, "On the statistical analysis of dirty pictures," *J. Royal Stat. Soc.*, vol. B48, no. 3, pp. 259-302, 1986.
- [25] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [26] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, no. 6, pp. 716-723, Dec. 1974.
- [27] J. Rissanen, "A universal prior for integers and estimation by minimum description length," *Ann. Stat.*, vol. 11, no. 2, pp. 416-431, 1983.
- [28] G. P. Robinson, A. C. Colchester, and L. D. Griffin, "Model-based recognition of anatomical objects from medical images," in *Proceedings of Information Processing in Medical Imaging*. Berlin, Germany: Springer-Verlag, 1993.
- [29] H. V. Poor, *An Introduction to Signal Detection and Estimation*. Berlin, Germany: Springer-Verlag, 1996.
- [30] R. Gray and L. Davisson, *Random Processes—A Mathematical Approach for Engineers*. Englewood Cliffs, NJ: Prentice-Hall, 1986.
- [31] Y. J. Zhang, "A survey on evaluation methods for image segmentation," *Pattern Recognit.*, vol. 29, no. 8, pp. 1335-1346, 1996.
- [32] K. Woods, Y. Wang, Z. Szabo, and M. T. Freedman, "Unsupervised tissue quantification and segmentation from 3-D MR brain images," in *Proc. Int. Signal Image Processing Conf.*, Las Vegas, NV, 1998.



Yu Wang (S'93-M'95) received the Ph.D. degree in electrical engineering from the University of Maryland Baltimore County, in 1995.

In 1996, he joined the Department of Electrical Engineering, The Catholic University of America, Washington, DC, where he is currently an Associate Professor. He is also an Adjunct Assistant Professor with the Johns Hopkins Medical Institutions, Baltimore, MD. His research interests include bioinformatics and bioimaging, information visualization, neural computation, image analysis, and fusion.

Dr. Wang was the recipient of the 1998 Department of Defense (CDMRP) Career Development Award and the 1999 Catholic University of America Outstanding Faculty Research Achievement Award.



Tülay Adalı (S'88-M'92) received the B.S. degree from the Middle East Technical University, Ankara, Turkey, in 1987, and the M.S. and Ph.D. degrees from North Carolina State University, Raleigh, in 1988 and 1992 respectively, all in electrical engineering.

In 1992, she joined the Department of Electrical Engineering, University of Maryland Baltimore County, Baltimore, where she is currently is an Associate Professor. Her research interests are in the areas of neural computation, adaptive signal processing, estimation theory, and their applications

in channel equalization, biomedical image analysis, time-series prediction, and optical communications.

Dr. Adalı is a member of the IEEE Neural Networks for Signal Processing Technical Committee and serves on the IEEE Signal Processing Conference Board. She has served on the Program Committees of a number of international conferences, including the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), the European Signal Processing Conference (EUSIPCO), and the IEEE International Workshop on Neural Networks for Signal Processing (NNSP). She is the Program co-chair for the IEEE NNSP'2000, general co-chair for the IEEE NNSP'2001, and the publicity chair for the IEEE ICASSP'2000. She was the recipient of a 1997 National Science Foundation CAREER Award.

Jianhua Xuan, photograph and biography not available at time of publication.



Zsolt Szabo received the M.D. degree from the University of Belgrade, Belgrade, Yugoslavia, and the Ph.D. degree from the University of Dusseldorf, Dusseldorf, Germany.

He received postgraduate training at the University of Dusseldorf, and the Johns Hopkins Medical Institutions, Baltimore, MD, as well as special training in positron emission tomography (PET) at the Nuclear Research Center, Julich, Germany. He is an Associate Professor of Radiology at the Johns Hopkins Medical Institutions. His research activities involve

PET/MR image registration, partial volume correction, investigation of serotonin and dopamine transporters in the brain, and angiotensin receptors in kidneys using PET.

Dr. Szabo is member of the Society of Nuclear Medicine and of the German Association of Nuclear Medicine.

NON-RIGID IMAGE REGISTRATION BY NEURAL COMPUTATIONS

Rujirutana Srikanchana Kelvin Woods Jianhua Xuan

Charles Nguyen Yue Wang

Department of Electrical Engineering and Computer Science
The Catholic University of America, Washington, DC 20064, USA

Abstract. This paper describes a neural computation based non-rigid registration methodology using multiple rigid transforms, in a piece-wise fashion, to model the registration process between images in a sequence. The registration methodology is a hybrid approach that combines registration without exact point correspondence via multi-object principal axes, and registration with point correspondence via polynomial transform. Neural computation is used to combine the derived individual principal axes solutions for each object in a committee machine formulation and to obtain the polynomial transform based on extracted control points using a multi layer perceptrons (MLP). Three examples are presented to demonstrate the techniques involved in the process. The first example uses four Gaussian clusters and focuses on the combination of the multiple transforms into a composite transform using finite mixture modeling techniques [14]. The next examples present the complete process for prostate cancer registration and breast sequence analysis respectively. To verify performance, the results are compared to non-neural based implementations and other existing registration methods.

INTRODUCTION

The estimation of transformational geometry from two point sets is an essential step to medical image fusion and computer vision [1, 2]. Medical diagnosis, for instance, often benefits from the complementary of the information in images of different modalities. The task is to recover a matrix representation requiring a set of correspondence matches between features in the two coordinate system [3]. Assume two 3-D data point sets $\{p_{iA}\}$ and $\{p_{iB}\}$; $i = 1, 2, \dots, N$ are related by

$$p_{iB} = R p_{iA} + T + N_i \quad (1)$$

where \mathbf{R} is a rotation matrix, \mathbf{T} is a translation vector, and \mathbf{N}_i is a noise vector. Given $\{\mathbf{p}_{iA}\}$ and $\{\mathbf{p}_{iB}\}$, Arun *et. al* present an algorithm for finding the least-squares solution of \mathbf{R} and \mathbf{T} , which is based on the decoupling translation and rotation and the singular value decomposition (SVD) of a 3×3 cross-covariance matrix [3].

The major limitation of the present method is twofold: (1) while feature matching methods can give quite accurate solutions, obtaining correct correspondences of features is a hard problem, especially in the cases of images acquired using different modalities or from inter-subjects; and (2) a rigidity assumption is heuristically imposed while those newly developed deformable matching methods can be computationally expensive and typically need good initial guesses to assure correct convergence [2, 4]. One popular method that is correspondenceless is the principal axes registration (PAR) [1], that is based on the relatively stable geometric properties of image features, i.e., the geometric information contained in these stable image features is often sufficient to determine the transformation between images [2]. Once again, it cannot handle the cases with non-rigid objects.

In this paper, we present a neural computation based non-rigid registration using piece-wise rigid transformation. The novel feature is to align two point sets without needing to establish explicit point correspondences, where the derivation is realized by minimizing the relative entropy between the two point distributions resulting in a maximum likelihood estimate of the transformation matrix. A committee machine approach is used for recovering the transformational geometry of the non-rigid structures. That is rather than using a single transformation matrix which gives rise to a large registration error, we attempt to interpolatively apply a mixture of transformations. By further generalizing PAR to a finite mixture registration (FMR) scheme, with a soft partitioning of the data set, the mixture is fit using expectation-maximization (EM) algorithm. We then applied a probabilistic adaptive principal components extraction (PAPEX) algorithm [10], to estimate the transformational of the orthogonal set of eigenvalues and eigenvectors of the auto-covariance matrix. By applying a committee machine to a non-rigid registration, using FMR as the experts and PAPEX as a gating function, we can acquire the registration based on a mixture of piece-wise transformations of the data set. Then the correspondences control points are obtained. As a final step, the warped image is obtaining using the neural network based non-linear mapping, to obtain the polynomial transform based on extracted control points using MLP.

THEORY AND METHOD

Suggested by information theory [5], we note that, since the control point sets in two images can be considered as two separate realizations of the same random source, we do not need to establish point correspondences to extract the transformation matrix. In other words, if we denote by $P_{\{\mathbf{p}_i\}}$ the distribution

of the control point set in an image, we have the simple relationship

$$P_{\{p_{jB}\}} = P_{\{Rp_{iA}+T\}} + \epsilon \quad (2)$$

where ϵ is the noise component. Since the probability distributions can be computed independently on each image without any need to establish feature correspondences, and given the two distributions of the control point sets in the two images, we can recover the transformation matrix in a simple fashion [2], as we now sketch.

For observation of the distributions, we can estimate R and T by minimizing the relative entropy (Kullback-Leibler distance) between $P_{\{p_{jB}\}}$ and $P_{\{Rp_{iA}+T\}}$. The least relative entropy estimator is then defined as

$$\arg \min_{R,T} D(P_{\{p_{jB}\}} || P_{\{Rp_{iA}+T\}}) \quad (3)$$

where D denotes the relative entropy measure. Following the same strategy to decoupling translation and rotation as in [3], we can define a new data point by $q_{iA} = p_{iA} - p_A^0$ and $q_{jB} = p_{jB} - p_B^0$, where p_A^0 and p_B^0 are the centroids of $\{p_{iA}\}$ and $\{p_{jB}\}$ respectively. Then the optimal geometric transformations, R and T , are defined as

$$R = U_B H U_A^t \text{ and } T = p_B^0 - R p_A^0 \quad (4)$$

where the superscript t denotes matrix transposition, U_A and U_B are 3×3 orthonormal matrices, and H is a 3×3 diagonal matrix with element $h_m = \sqrt{\lambda_{mB}/\lambda_{mA}}$. Note that the transformation U consists of the orthonormal set of eigenvectors and h_m is the squared root of the eigenvalues λ_m of the auto-covariance matrix C for $m = 1, 2, 3$ and for $\{p_{iA}\}$ and $\{p_{jB}\}$, respectively.

However, because of its global linearity, the application of PAR is necessarily somewhat limited [6]. An alternative paradigm is to model a multimodal control point set with a collection of local linear models. The method is a two-stage procedure: a soft partitioning of the data set followed by estimation of the principal axes within each partition [7]. Recently there has been considerable success in using standard finite normal mixture (SFNM) to model the distribution of a multimodal data set and the association of a SFNM distribution with PAR offers the possibility of being able to register two images through a mixture of probabilistic principal axes transformations [7]. Assume that there are K_0 control point clusters, where each control point cluster defines a transformation $\{R_k, T_k\}$. Thus for a point p_{nA} , its new locations corresponding to each of the transformations are $p_{nk} = R_k p_{nA} + T_k$ for $k = 1, \dots, K_0$. Further assume that the control point set defines a SFNM distribution

$$f(p_i) = \sum_{k=1}^{K_0} \pi_k g(p_i | \mu_k, C_k) \quad (5)$$

where g is the Gaussian kernel with mean vector μ_k and auto-covariance matrix C_k , and π_k is the mixing factor proportional to the number of control

points in cluster k . For each of the control point sets $\{p_{iA}\}$ and $\{p_{iB}\}$, the mixture is fit using the expectation-maximization (EM) algorithm. The E step involves assigning to the linear models responsibilities from the control points; the M step involves re-estimating the parameters of the linear models in the light of this assignment [7].

Thus the statistical membership of point p_{nA} belonging to each of the control point clusters can be derived by

$$z_{nk} = P(\mathbf{R}_k, \mathbf{T}_k | p_{nA}) = \frac{\pi_{kA} g(p_{nA} | \mu_{kA}, \mathbf{C}_{kA})}{f(p_{nA})} \quad (6)$$

i.e., the posterior probability of $\{\mathbf{R}_k, \mathbf{T}_k\}$ given p_{nA} . Thus, we can define the FMR transformation as

$$\mathbf{p}_n = \sum_{k=1}^{K_0} z_{nk} \mathbf{p}_{nk} = \sum_{k=1}^{K_0} \frac{\pi_{kA} g(p_{nA} | \mu_{kA}, \mathbf{C}_{kA})}{f(p_{nA})} (\mathbf{R}_k \mathbf{p}_{nA} + \mathbf{T}_k) \quad (7)$$

where $\{\mathbf{R}_k, \mathbf{T}_k\}$ is determined based on $f(p_i) = \sum_{k=1}^{K_0} \pi_{kA} g(p_i | \mu_{kA}, \mathbf{C}_{kA})$ that we have estimated in the previous step using the EM algorithm. Note that now we do need the correspondences between the two control point clusters, and these correspondences may be found, after a global PAR is initially performed, by using a site model supported approach or a dual-step EM algorithm to unify the tasks of estimating transformation geometry and identifying cluster-correspondence matches [4]. This philosophy for recovering transformational geometry of the non-rigid objects is similar in spirit to the *divide-and-conquer* principle [6], under which the relative entropy between the two point sets reaches its minimum

$$\arg \min_{\mathbf{R}_k, \mathbf{T}_k} D(P_{\{p_{iB}\}} || P_{\{\sum_{k=1}^{K_0} z_{ik} (\mathbf{R}_k \mathbf{p}_{iA} + \mathbf{T}_k)\}}) \quad (8)$$

both globally and locally.

Based on a mixture of probabilistic principal axes transformations, the next section describes a neural computation using a committee machine approach for which a complex computational task is solved by dividing it into a number of computationally simple tasks and then combining the solutions to those tasks. Basically, it fuses knowledge acquired by experts to arrive at an overall decision that is supposedly superior to that attainable by any one of them acting alone.

NEURAL COMPUTATION

A neural network interpretation of the EM algorithm is given in [8]. Because of its reputation of being slow in which new information acquired in the expectation step is not used immediately, on-line versions of the EM algorithm are proposed for large-scale sequential learning. Thus, we adopt a fully unsupervised and incremental stochastic learning algorithm. The

scheme provides winner-takes-in probability (Bayesian "soft") splits of the control points, hence allowing the data to contribute simultaneously to multiple clusters which results in

E - Step

$$z_{(i+1)k} = \frac{\pi_k^{(i)} g(\mathbf{p}_{i+1} | \mu_k^{(i)}, \mathbf{C}_k^{(i)})}{f(\mathbf{p}_{i+1} | \pi_k^{(i)}, \mu_k^{(i)}, \mathbf{C}_k^{(i)})} \quad (9)$$

M - Step

$$\mu_k^{(i+1)} = \mu_k^{(i)} + a(i)(\mathbf{p}_{i+1} - \mu_k^{(i)})z_{(i+1)k}, \quad (10)$$

$$\mathbf{C}_k^{(i+1)} = \mathbf{C}_k^{(i)} + b(i)[(\mathbf{p}_{i+1} - \mu_k^{(i)})(\mathbf{p}_{i+1} - \mu_k^{(i)})^T - \mathbf{C}_k^{(i)}]z_{(i+1)k}, \quad (11)$$

$$\pi_k^{(i+1)} = \frac{i}{i+1} \pi_k^{(i)} + \frac{1}{i+1} z_{(i+1)k} \quad (12)$$

for $k = 1, \dots, K_0$ and for $\{\mathbf{p}_{iA}\}$ and $\{\mathbf{p}_{iB}\}$, respectively, where $a(i)$ and $b(i)$ are introduced as the learning rates, two sequences converging to zero, ensuring unbiased estimates after convergence. This procedure is termed as neural computation of the EM algorithm, where at each complete cycle of the algorithm, we first use "old" set of parameter values to determine the posterior probabilities $z_{(i+1)k}$. These posterior probabilities are then used to obtain "new" values $\pi_k^{(i+1)}, \mu_k^{(i+1)}, \mathbf{C}_k^{(i+1)}$. The algorithm cycles back and forth until the value of relative entropy between the data histogram and mixture model reaches its minimum

$$\arg \min_{\pi_k, \mu_k, \mathbf{C}_k} D(P_{\{\mathbf{p}_i\}} || f(\mathbf{p}_i)) \quad (13)$$

for $\{\mathbf{p}_{iA}\}$ and $\{\mathbf{p}_{iB}\}$, respectively.

With a soft partitioning of the data set using Eqs. (9-12), control points will now effectively belong to more than one cluster spatially. Thus, the effective input values are $\mathbf{p}_{ik} = z_{ik}(\mathbf{p}_i - \mu_k)$ for an independent registration transformation k in the committee machine [9]. We then extend our adaptive principal components extraction (APEX) algorithm to a probabilistic version, i.e., PAPEX [10], to determine \mathbf{U}_k for $\{\mathbf{p}_{iA}\}$ and $\{\mathbf{p}_{iB}\}$, respectively, summarized as follows.

1. Initialize the feedforward weight vector \mathbf{u}_{mk} for $m = 1, 2, 3$, and the feedback weight vector a_{mk} to small random values for $m = 2, 3$, at time $i = 1$. Assign a small positive value to the learning rate parameter η .
2. Set $m = 1$, and for $i = 1, 2, \dots$, compute

$$y_{1k}(i) = \mathbf{u}_{1k}^T(i) z_{ik}(\mathbf{p}_i - \mu_k) \quad (14)$$

$$\mathbf{u}_{1k}(i+1) = \mathbf{u}_{1k}(i) + \eta[y_{1k}(i) z_{ik}(\mathbf{p}_i - \mu_k) - y_{1k}^2(i) \mathbf{u}_{1k}(i)] \quad (15)$$

For large i we have $\mathbf{u}_{1k}(i) \rightarrow \mathbf{u}_{1k}$, where \mathbf{u}_{1k} is the eigenvector associated with the largest eigenvalue of the cluster k , and $\lambda_{1k} = \frac{1}{N} \sum_{i=1}^N y_{1k}^2(i)$.

3. Set $m = 2$, and for $i = 1, 2, \dots$, compute

$$\mathbf{y}_{(m-1)k}(i) = [y_{1k}(i), y_{2k}(i), \dots, y_{(m-1)k}(i)]^T \quad (16)$$

$$\mathbf{y}_{mk}(i) = \mathbf{u}_{mk}^T(i) \mathbf{z}_{ik}(\mathbf{p}_i - \boldsymbol{\mu}_k) + \mathbf{a}_{mk}^T(i) \mathbf{y}_{(m-1)k}(i) \quad (17)$$

$$\mathbf{u}_{mk}(i+1) = \mathbf{u}_{mk}(i) + \eta[y_{mk}(i) \mathbf{z}_{ik}(\mathbf{p}_i - \boldsymbol{\mu}_k) - y_{mk}^2(i) \mathbf{u}_{mk}(i)] \quad (18)$$

$$\mathbf{a}_{mk}(i+1) = \mathbf{a}_{mk}(i) - \eta[y_{mk}(i) \mathbf{y}_{(m-1)k}(i) + y_{mk}^2(i) \mathbf{a}_{mk}(i)] \quad (19)$$

For large i we have $\mathbf{u}_{2k}(i) \rightarrow \mathbf{u}_{2k}$, where \mathbf{u}_{2k} is the eigenvector associated with the second largest eigenvalue of the cluster k , and $\lambda_{2k} = \frac{1}{N} \sum_{i=1}^N y_{2k}^2(i)$.

4. Set $m = 3$, go to step 3. For large i we have $\mathbf{u}_{3k}(i) \rightarrow \mathbf{u}_{3k}$, where \mathbf{u}_{3k} is the eigenvector associated with the third largest eigenvalue of the cluster k , and $\lambda_{3k} = \frac{1}{N} \sum_{i=1}^N y_{3k}^2(i)$.

The neural computational of a committee machine can be achieved by distributing the learning tasks among a number of experts, which in turn partitioning the input space into a set of subspaces. The experts are in theory performing supervised learning in that there individual outputs are combined to model the desired response. There is, however, a sense in which the experts are also performing self-organized learning; that is, they self organize to find a good partitioning of the input space so that each expert does well at modeling its own subspace, and as a whole group they model the input space well.

Consider the modular network nature of Eq. (7), it is a mixture of local expert model, in which the individual responses of the experts are nonlinearly combined by means of a single gating network. It is assume here that the different experts work best in different regions of the input space in accordance with the probabilistic generative model. In our case, the effective input values are $\mathbf{p}_{ik} = \mathbf{z}_{ik}(\mathbf{p}_i - \boldsymbol{\mu}_k)$ for an independent registration transformation k in the networks. A committee machine consists of k supervised modules, with a soft partitioning of the data set using EM algorithm called experts, and an integrating unit called a gating network that performs the function PAPEX to determine the transformation of the orthogonal set of eigenvalues and eigenvectors of the auto-covariance matrix among the expert networks. The output of our committee machine is a transformational matrix of image pair. Based on those correspondences between the two images, the control points are obtained. The final step is to calculate the polynomial transformation using piece-wise interpolation. In our case for a given corresponding control points pair of the image, we considered neural network based MLP to acquired the polynomial transform. The final result shows the warped image that has been corrected by applied our neural computation to correct most of the scale different between the images.

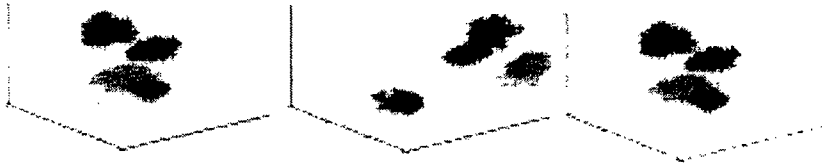


Figure 1: A mixture Gaussian data set (left: Gaussian model; middle: FMR registration result; right: inverse FMR result)

RESULTS AND DISCUSSIONS

To evaluate the effectiveness of our algorithm, three simulations were performed. The first simulation explored the performance of FMR under controlled conditions while the other two simulations considered the complete processing chain applied to real world problems. Next, each of these simulations is discussed.

The first simulation is a mixture of 600 data points generated from four Gaussian clusters in 3-D space. Figure 1(left) is the original image; the global view of the data clearly suggests the presence of four distinct clusters within the data. We considered three of the four Gaussian clusters as the control objects and the remaining object as a noncontrol object. Each of the clusters were translated and rotated by different amounts to simulate a non-global transform. With a mixture of all the three local transforms, we acquired a non-rigid transform of a non-control object based on a committee machine approach as shown in Figure 1(middle). To show the robustness of our propose method, we applied our algorithm to the transformed image as an inverse process resulting in Figure 1(right). The performance metric is mean square error (MSE) of the non-control object. Several runs of this example were conducted to average out randomness of the data points. The results show a MSE of 7.6% between the final image and original. Thus, showing the usefulness of this approach.

Next, we extended our algorithm to a 3-D prostate registration problem. Based on our previous work [14], this 3-D model contains a precise probabilistic map of prostate tumor distribution and the corresponding anatomic structure of a prostate. The goal of this example is to register the tumor inside a prostate, with a reference model using prostate anatomic structure as a control objects. We denote the reference model as shown in Figure 2(left) and the float model with non-control object as shown in Figure 2(middle). We find object to object corresponding and let tumor in the float model be a non-control object. After we applied our hybrid committee machine algorithm to the reference and the float model, based on piece-wise registration, each control points are effectively belong to more than one cluster hence allowing the data to contribute simultaneously to multiple clusters. Figure 2(right) shows the result using neural computation to register tumor in a float model to a reference model.

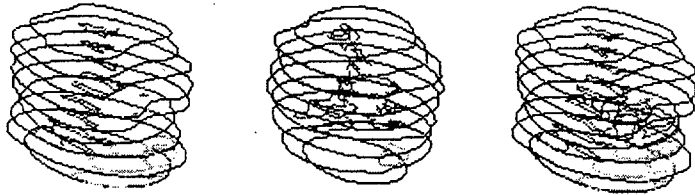


Figure 2: Prostate Model used in registration (left: reference model; middle: float image; right: registration cancer class of float image to reference image using FMR method)

As a final example, we applied our algorithm to a temporal sequence of mammograms of a single patient [15]. In this example, the committee machine is used to obtain an initial registration using multiple extracted objects (skinline, dense tissue regions) in a finite mixture scheme. Then MLP was used to determine the coefficients of a polynomial transform using extracted vertical and horizontal cross points of elongated structures as control points. Previously, thin-plate spline (TPS) was used to determine the transform coefficients [13]. As a comparison, we consider both results here. Figure 3 shows the raw sequence and Figure 4 shows the resulting warped image for MLP (left) and TPS (right). From visual inspection we see that by using MLP the most of the scale different between images has been corrected. While the TPS distorts the image, this distortion can be attributed to control point selection and correspondence. The MLP better adapts to the error present in the control points thus yielding a smoother result.

CONCLUSION

In this paper, we presented the theoretical concepts and methods of a neural computation based non-rigid registration algorithm. The approach uses

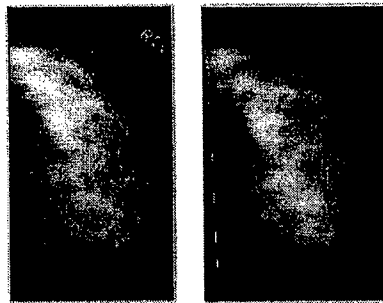


Figure 3: A pair of real mammograms taken over a period of time from the same patient.

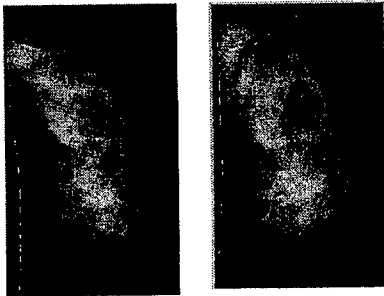


Figure 4: The result of warped current image to previous image (left: MLP; right: TPS).

a committee machine to recover the total transformational geometry of the non-rigid object using multiple rigid transforms combined together in a finite mixture registration scheme. Finite mixture transform combination is a novel technique to combine multiple transforms that are contained in a single image. Other than local transforms no other method combines multiple transforms together. In addition finite mixture combinations, yields a lower MSE than the local transform and results in a smooth image while local transforms yield an image containing discontinues on transform boundaries. In addition, the registration obtained in the committee machine is fine tuned using a non-linear transform generated by a MLP network using extracted control points.

We applied this algorithm to prostate and breast registration problems with reasonable results as shown in the previous figures. Some distortion can be seen in the final warped images because of the error in control point selection and correspondence. Improvement in this portion should decrease distortion and yield a smoother looking image. Using neural networks in this problem has increased the generality of this approach by allowing the algorithm to adjust performance as imaging condition change.

ACKNOWLEDGMENTS

This work was supported by the US Army Medical Research and Materiel Command under Grant DAMD17-00-0195. The authors would like to thank the experts whose name are presented in our early papers for their valuable time and great assistance.

REFERENCES

- [1] M. Moshfeghi and H. Rusinek, "Three-dimensional registration of multimodality medical images using the principal axes techniques," *Philips J. Res.*, vol. 47, no.2, pp. 81-97, 1992.

- [2] V. Govindu and C. Shekhar, "Alignment using distributions of local geometric properties," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, no.10, pp. 1031-1043, October 1999.
- [3] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-Squares Fitting of Two 3-D Point Sets," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, no.5, pp. 698-700, September 1987.
- [4] A. D.J. Cross and E. R. Hancock, "Graph matching with a dual-step EM algorithm," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no.11, pp. 1236-1253, November 1998.
- [5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.
- [6] G. E. Hinton, P. Dayan, and M. Revow, "Modeling the manifolds of images of handwritten digits," *IEEE Trans. Neural Nets*, vol. 8, no.1, pp. 65-74, January 1997.
- [7] Y. Wang, L. Luo, M. T. Freedman, and S. Y. Kung, "Probabilistic principal component subspaces: A hierarchical finite mixture model for data visualization," *IEEE Trans. Neural Nets*, vol. 11, no. 3, pp. 625-636, May 2000.
- [8] L. Perlovsky and M. McManus, "Maximum likelihood neural networks for sensor fusion and adaptive classification," *Neural Networks*, vol. 4, pp. 89-102, 1991.
- [9] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed., Prentice-Hall, Inc., Upper Saddle River, New Jersey, 1999.
- [10] S. Y. Kung, *Principal Component Neural Networks*, New York: Wiley, 1996.
- [11] M. I. Jordan and R. A. Jacobs, "Hierarchical mixture of experts and the EM algorithm," *Neural Computation*, vol. 6, pp. 181-214, 1994.
- [12] D. M. Titterton, A. F. M. Smith, and U. E. Markov, *Statistical analysis of finite mixture distributions*. New York: John Wiley, 1985.
- [13] K. Woods, et. al., "Patient Site Model Supported Change Detection," *Proceedings of SPIE Medical Imaging: 2000*, vol.1, no. 24 , pp. 1095-1106, 2000.
- [14] J. Lu, R. Srikanthana, M. McClain, Y. Wang, J. Xuan, I. A. Sesterhenn, M. T. Freedman, and S. K. Mun "A Statistical Volumetric Model for Characterization and Visualization of Prostate Cancer," *Proc. SPIE Medical Imaging*, vol. 3976, pp. 142-153, 2000.
- [15] K. Woods, *Image Guided Diagnosis through Change Detection in Image Sequences*, Doctoral Dissertation, The Catholic University of America, 2000.

A Multiple Circular Path Neural Network Architecture for Detection of Mammographic Masses

Shih-Chung B. Lo^a, Huai Li^b, Akira Hasegawa^c, Yue J. Wang^d, Matthew T. Freedman^a,
and Seong K. Mun^a

^aISIS Center, Radiology Department, Georgetown University Medical Center, Washington, DC.

^bOdyssey Technologies Inc., Jessup, Maryland.

^cR2 Technology, Inc. Los Altos, California

^dDepartment of Electrical Engineering and Computer Sciences, Catholic University of America, Washington, DC.

Li, Hasegawa, and Wang were with the ISIS Center, Radiology Department, Georgetown University Medical Center where they worked with the other authors to contact this study.

Please send correspondence to:

Shih-Chung B. Lo, Ph.D.

Professor and Chief Scientist of ISIS Center

ISIS Center, Radiology Department

Georgetown University Medical Center

2115 Wisconsin Avenue, Suite 603, N.W.

Washington, DC, 20007

Abstract

In the clinical course of detecting masses, mammographers usually evaluate the surrounding background of a radiodense when breast cancer is suspected. In this study, we adapted this fundamental concept and computed features of the suspicious region in radial sections. These features were then arranged by circular convolution processes within a neural network, which led to an improvement in detecting mammographic masses.

In this study, randomly selected mammograms were processed by morphological enhancement techniques. Radiodense areas were isolated and were delineated using a region growing algorithm with a valley blocking technique. The boundary of each region of interest was then divided into 36 sectors using 36 equi-angle dividers radiated from the center of the area. Four features at each section were computed: (1) the radius, (2) the normal angle of the boundary, (3) the average gradient along the radial direction, and (4) the gray value difference (i.e., contrast) along the radial direction. Hence, 144 computed features (i.e., 4 features per sector for 36 sectors) were used as input values for the newly invented multiple circular path neural network (MCPNN). The neural network is constructed to emphasize on the correlation information associated with the feature interactions within the angle and between adjacent angles.

We have tested this approach on our research database consisting of 91 mammograms. The over-all performance in the detection of masses was 0.78-0.80 for the areas (A_z) under the ROC curves using the conventional neural network. However, the performance was improved to A_z values of 0.84-0.89 using the multiple circular path neural network.

1. Introduction

It is well known that effective treatment of breast cancer calls for early detection of cancerous lesions (e.g., clustered microcalcifications and masses associated with malignant cellular processes) [1-3]. Breast masses appear as areas of increased density on mammograms. It is particularly difficult for radiologists to detect and analyze a suspected area where a mass is overlapped with dense breast tissue. These masses are more readily seen as time progresses, but the further the tumor has progressed, the lower the possibility of a successful treatment. Therefore, increasing the chances of early breast cancer detection in improving today's clinical system is of vital importance to breast cancer patients.

Several research groups have developed computer algorithms for automated detection of mammographic masses [4-8]. Investigators also attempted to classify the malignant or benign nature of the detected tumors [5]. The results of these detection programs indicate that a high true-positive (TP) rate can be obtained at the expense of 2 or 3 false-positive (FP) detections per mammogram. This FP rate is unacceptably high for mass detection in clinical practice. Mammographically, a multiplicity (more than two) of similar benign-appearing breast lesions argues strongly for benignity [9-12] and, indeed, the more masses that are identified, the less chance that they represent cancer [12]. If the computer indicates multiple detections on each mammogram, the radiologist has to seek out the one mass that has mammographic features that differ from the others. The significant lesion may be missed due to the multiplicity of possible lesions. We therefore believe that a more useful and fundamental approach to computer-aided diagnosis (CADx) of masses is to devise computer programs to analyze features of a suspected mass [13,14], which are detected by the radiologist, and provide feature measures and estimates of the likelihood of malignancy by comparing the computer's database. The computer therefore serves as a second opinion and also provides a reproducible and objective evaluation of the mass. With this aid, the radiologist may also increase his/her sensitivity by lowering the threshold of suspicion, while maintaining the overall specificity and reading efficiency.

2. Clinical Background of Breast Lesions and Technical Approach in Mass Detection

2.1 Description of clinical background

Most commonly, breast cancer presents as a mass. The same lesion shows a somewhat different picture from one projection to the other. Difficulties in mass detection also vary with the underlying breast parenchyma. In the fatty breast, masses are generally easy to detect. With the dense breast, mass detection is more difficult and auxiliary signs aid this detection. Breasts can contain one, several, or many masses. When there is one mass, the decision process is based on its size, shape, and margins. The larger the mass is and the less well-defined its margins, the greater the chance of cancer. When there are several masses, one looks at each, trying to determine whether any has features to suggest cancer (poorly defined, spiculate, unusually radiodense for size) and one also looks to see whether any mass is different in appearance from the others. Multiple small, well-defined, similar masses presenting bilaterally are all likely to be benign. The greater the asymmetry, size, lack of circularity, edge unsharpness, and radiodensity, the more suspicious. In this study, we used several computational features (see section 3.2) highly associated with four major features of breast masses routinely used in clinical reading:

Density - Malignant lesions tend to have greater radiographic density due to high attenuation and less compressibility of cancer than normal tissue. Radiolucent lesions are typically benign and the diagnosis can be made from the mammogram.

Size - If the lesion has morphological features suggesting malignancy, it should be considered suspicious regardless of the size. Isolated masses with non-cystic densities greater than 8 mm in diameter can be malignant. In general, the larger a lesion, the more suspicious it is.

Shape - The more irregular the shape of a lesion, the more likely the possibility of malignancy. Lesions tend to

be round, ovoid and/or lobulated. Small and frequent lobulations are suspicious. Lesions in the lateral aspect of the breast near the edge of the parenchyma with a reniform shape and a hilar indentation or notch usually represent a benign intramammary lymph node. Breast carcinoma hidden in the dense tissues can cause parenchymal retraction, which possess different shapes.

Margins -The margins of the lesion should be carefully evaluated for areas of spiculation, stellate patterns or ill-defined regions. Most breast cancers have ill-defined margins secondary to tumor infiltration and associated fibrosis. The appearance of spiculations and a more diffuse stellate pattern are almost pathognomonic for cancer. Lesions with sharply defined margins have a high likelihood of being benign; however, up to 7% of malignant lesions can be well circumscribed.

These are known clinical features and have been adapted in "Breast Imaging - Reporting and Data System" (BI-RAD) [15] of the American College of Radiology (ACR).

2.2. Technical approach for detection of mammographic masses

In this study, our goal was to extract clinically suspicious lesions. The differentiation of benign and malignant status was beyond the scope of this work. Hence, we will only provide methods in extracting potential lesions from glandular tissue in the following sections. (Note that lesions can be overlapped with dense breast parenchyma.) The study was conducted with the following steps: (1) use background correction method and morphological operations to extract radio-opaque areas, (2) delineate the boundary of the areas, (3) compute the features and texture of the masses with emphasis on the boundary, and (4) design training strategy using neural networks as classifiers for the recognition of mass features. An overall detection scheme of the study framework is shown in Figure 1.

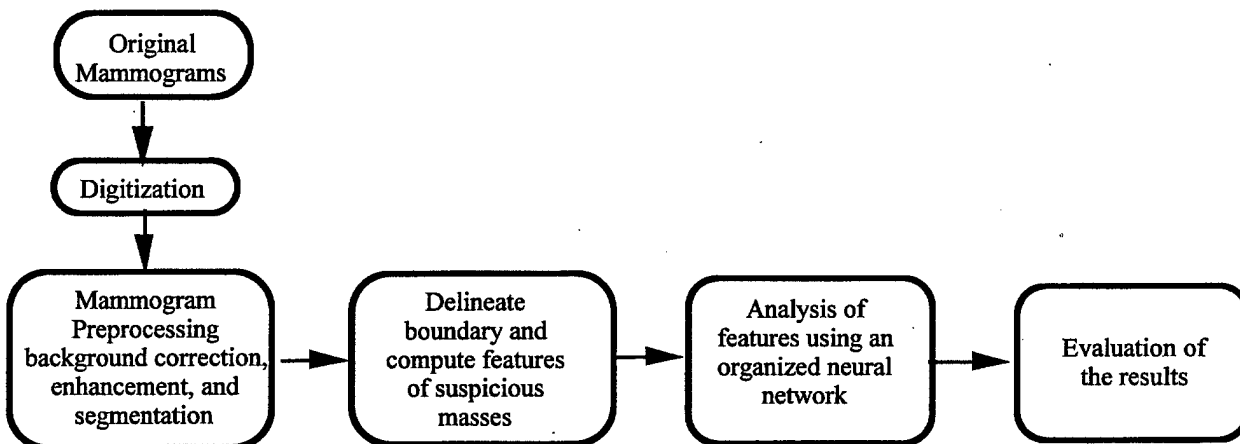


Figure 1. A system flowchart for the detection of masses in this study.

3. Development of Technical Methods

3.1. Preprocessing for image consistency and mass enhancement using morphological operations

One of the main difficulties in automatic mass-detection is that mammographic masses are often overlapped with breast tissues. In such cases, it is necessary to remove bright background caused by breast tissues but to keep mass-signals. For this purpose, background correction is an indispensable technique for mass detection.

The theory of mathematical morphology is powerful in analyzing and describing geometrical relations. Essentially it is a formalization of intuitive concepts such as size or shape. The two basic morphological operations are "erosion" and "dilation," which are consistently defined for binary and gray-scale images. Using

these two basic operations, two other basic and important operators, "opening" and "closing", can be defined as follows:

$$\text{opening:} \quad X_B \equiv (X \ominus B) \oplus B, \quad \dots(1)$$

$$\text{closing:} \quad X^B \equiv (X \oplus B) \ominus B, \quad \dots(2)$$

where X indicates the original image, B represents the structuring element, and \oplus and \ominus indicate the operations "dilation" and "erosion," respectively. Based on the "opening" operation, we have developed an operation for background correction. The operation is represented by

$$X - X_B = X - (X \ominus B) \oplus B. \quad \dots(3)$$

This equation represents the subtraction of the image processed by the operator "opening" from the original image.

Figure 2 shows the effect of the operation represented by equation (3): (A) illustrates a structuring element, (B) shows the original signal (gray line) and the processed signal (black line) by "opening", and (C) denotes the final output signal of the morphological operation. The final profile in (C) was obtained by subtracting the black profile signal from the gray profile signals in (B). Note that the detected peak signals were not affected by the operation. Hence the mass signals detected by the operation retain their original shapes.

As can be seen in this graph, the size of the detected peak significantly depends on the size of the structuring element. All peaks, which are smaller than the structuring element, can be detected. In our mass detection process, a 52 pixel-diameter structuring element will be used to detect masses whose sizes are less than 52 pixels in diameter. An object with a diameter of 52 pixels in a 512×625 pixel reduced image occupies 250 pixels in its original digitized image, and its real size is expected to be about 2.5 cm.

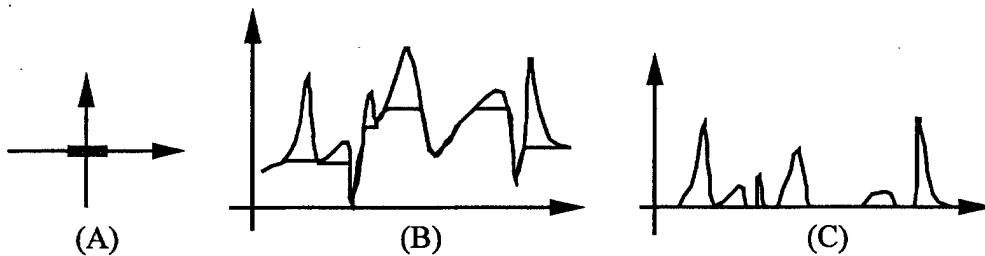


Figure 2. An example of the morphological operation: (A) structuring element, (B) original signal (gray line) and signal after opening (black line), and (C) output signal of the morphological filtering.

3.2. Feature extraction of masses

Feature extraction methods have played essential roles in many pattern recognition tasks. Once the features associated with an image pattern are extracted accurately, they can be used to distinguish one class of patterns from the others. Recently, many investigators have found that the multilayer perceptron neural network using the error back propagation training technique is a very powerful tool to serve as an analyzer (or classifier). Recently, the back propagation neural network (BPNN) for classification of features has widely been used in the field of computer-aided diagnosis [16-20].

The success of using an analyzer for a pattern recognition task would rely on two issues: (a) selected features that could describe discrepancy between patterns and (b) accuracy of the feature computation. Should either one fail, no analyzer or classifier would be able to achieve the expected performance. By analyzing many clinical samples of various sizes of masses, we found that the peripheral portion of the mass plays an important role for mammographers to make a diagnosis. The mammographer usually evaluates the surrounding background of a radiodense area when breast cancer is suspected.

We, therefore, performed boundary detection of the suspected masses on the morphologically enhanced mammogram. A region growing with valley blocking technique was employed to delineate all the suspected areas. Then, the boundary was divided into 36 sectors (i.e., 10° per sector) using 36 equi-angle dividers radiated from the center of suspicious area. The following features were computed within each 10° sector of the area:

- (a) "l" - the length from the center of mass to the shortest boundary segment.
- (b) "a" - the normal angle of the boundary segment (or the value of $\cos(a)$).
- (c) "g" - the average gradient of gray value on the segment along the radial direction.
Technically speaking, this set of gradient values may also serve as a fuzzy system for the input layer in the neural network to be described.
- (d) "c" - the gray value difference (i.e., contrast) along the radial direction. Averaged gray value (h_i) calculated from the mass area located at $l/3$ inside the boundary and the average background value (b_0) calculated from the peripheral area near $l/3$ outside of the suspicious area).

Hence, a total of 144 computed features (4 features/sector for 36 sectors) can be used as input values for the analysis of suspicious areas. The relationship between the computed features and BI-RADS descriptors are discussed below:

- (1) Mass Size -
The 36 "l" values would provide sufficient data for the neural network to determine the size.
- (2) Mass Shape (round, oval, lobulated, or irregular) -
The 36 "l" and 36 "a" values could approximate the shape of a mass.
- (3) Mass Margin (circumscribed, microlobulated, obscured, ill-defined, or spiculate) -
The 36 "g" and 36 "l" values should be able to describe the characteristics of the mass margin.
- (4) Mass Density (fat-containing, low density, isodense, or highly dense) -
The 36 "c" and 36 "g" values would be able to describe the density of the mass.

In short, the selected features are greatly associated with the main mass descriptors indicated in the BI-RADS. The reason for using 36 values for each nominated feature is four-fold: (a) mass boundary varies, it is difficult to describe an image pattern using a single value; (b) due to the general shape of the masses, the features of masses can be easily analyzed by the polar coordinate system; (c) in case some features are inaccurately computed in several directions due to the structure noises, such as the breast slender lines, there may still exist a sufficient number of correct features; (d) generally more accurate results can be produced by using subdivided parameters rather than using global parameters in a pattern recognition task. Other computational features (e.g., difference entropy [8] and other higher order features) are eligible but require further investigation.

3.3. The neural network structures specifically designed for the extracted boundary features

(A) Multiple paths with circular networking to instruct the neural network in analyzing sector features

We designed several neural network connections between the input and the first hidden layers as shown in Figure 3. Figure 3(A), (B), and (C) illustrate the full connection, a self correlation (SC) networking, and a neighborhood correlation (NC) networking, respectively. Note that the input and hidden nodes should be completely matched when combining more than one path in the study. In this case, the correlation layers only

function as branch connections between input and hidden layers. When using NC paths, networking engagement within multiple sectors (e.g., 20°, 30°, 40°, and 50° of the neighborhood correlation) can be grouped. The method of using the multiple correlation connections was motivated by our two-dimensional convolution neural network (2-D CNN) research experience where we found that more than 10 multiple convolution kernels were necessary to archive an outstanding neural network performance in the detection of lung nodules and microcalcifications [17].

Compared to 2-D CNN systems, the required computation using 1-D input features (i.e., 144) is relatively small. The combination of the networking paths described earlier for MCPNN was implemented using C programming language. The internal computation algorithm used in the MCPNN shares the same convolution process as that in the 2-D CNN [17]. One additional training method using flipping invariance for 1-D convolution kernels was employed and is described in the section 3.3.(B).

The fully connected neural network is a standard back propagation neural network. The signals of the fully connected neural network join the other two network processes (SC and NC paths) at the single node of the output layer. The signal received at the output node is scaled between 0 and 1. During the training, 0 and 1 were assigned at the output node to perform back propagation computation for a non-mass and a mass, respectively. The back propagation was computed in such a way that the computed incremental errors (see equations (9) and (10)) were retraced into three independent network paths (full-connected, SC, and NC paths). Besides the output layer, the SC and NC signals were independently arranged and are processed through two one-dimensional convolution processes in the forward propagation. The learning algorithms for all three paths were based on the backpropagation training method.

Let $N^0(n, s)$ represents input signal at the node n and sector s . The signal received at each node on the first hidden layer of the SC path is

$$N_{sc}^1(s) = \left(\sum_n N^0(n, s) \times W_{sc}(n) \right) + b_{sc}(s), \quad (4)$$

where $b_{sc}(s)$ represents the bias in s th sector. The signal gets into each node on the first hidden layer of the NC path is

$$N_{nc}^1(s) = \left(\sum_{s=-s1}^{s1} \sum_n N^0(n, s) \times W_{nc}(n, s) \right) + b_{nc}(s), \quad (5)$$

where $b(s)$ represents the bias in s th sector and $s1$ is 2 to cover -20 degree to 20 degrees of the fan. The signals in other hidden layers in each path are processed the same as the standard fully connected neural network. The output signal was collected from the last hidden layer and is given by,

$$O = S(N_p^l(n) \cdot W_p^l(n)), \quad (6)$$

where l denotes the hidden layer, p denotes the path, and $S(z)$ is a sigmoid function given by

$$S(z) = \frac{2}{1 + \exp(-z)} \quad (7)$$

The sigmoid function would produce modulated values ranging from 0 to 1.

Let the t -th change of the weight be $\Delta W_p^l(n)$ and the t -th change of the bias be $\Delta b_p^l(t)$. The error function is defined as

$$E = \frac{1}{2} (T - O)^2 \quad (8)$$

where T and O denote the desired output value and the actual output value, respectively when the input nodes $N^0(n,s)$, are entered in the network. In this model, the error back propagation algorithm, which updates the kernel weights, was given below:

$$\Delta W_p^l[t+1] = \eta \left(\sum \delta_p^{l+1}(n,s) \cdot O_p^l(n,s) \right) + \alpha \Delta W_p^l[t] \quad (9)$$

$$\Delta b_p^l[t+1] = \eta \sum \delta_p^l(n,s) + \alpha \Delta b_p^l[t] \quad (10)$$

$$\delta_p^l(n,s) = N_p^L(n,s) \left(\delta_p^{l+1}(n,s) \cdot W_p^l(n,s) \right) \quad (11)$$

where $s = 0$ when $l \neq 0$. In the case of the last layer,

$$\delta_p^L = S'(N_p^L(n)) (T - O) \quad (12)$$

where $S'(z)$, η , α , and T denote the derivative of $S(z)$, the learning rate, the weighting factor contributed by the momentum term, and the desired output image, respectively.

During the training, we added an isotropic constraint to the weights in the 1-D convolution kernels and so that

$$W_p^0(n, -s) = W_p^0(n, s) \quad (13)$$

where p is not the fully connected path. These additional constraints were used to induce the kernels functioning as correlation processing filters and could facilitate the algorithm in searching for an appropriate linear filter.

(B) Training methods and the utilization of characteristics of flipping invariance of the features

Because we used the circular paths, there were no starting and ending sectors. The forward and back propagation computation can be started from any sector. Since the mass characteristics of the flipped patch remained the same, we flipped each patch in the training set and kept the same numerical value for the target output.

Since we designed a 10° increment for each rotation, each SC or NC networking would need to process through 36 times for the computed feature set for each image patch. To simplify this network computation, we shifted one small set (4 nodes) on the input layer a time to conduct the circular convolution process with the SC and NC kernels. By reversing the sequence of the sector, we can train the flipped version of the suspicious masses. Hence, the characteristics of flipping invariance literally increase the number of the training set by a factor of 2.

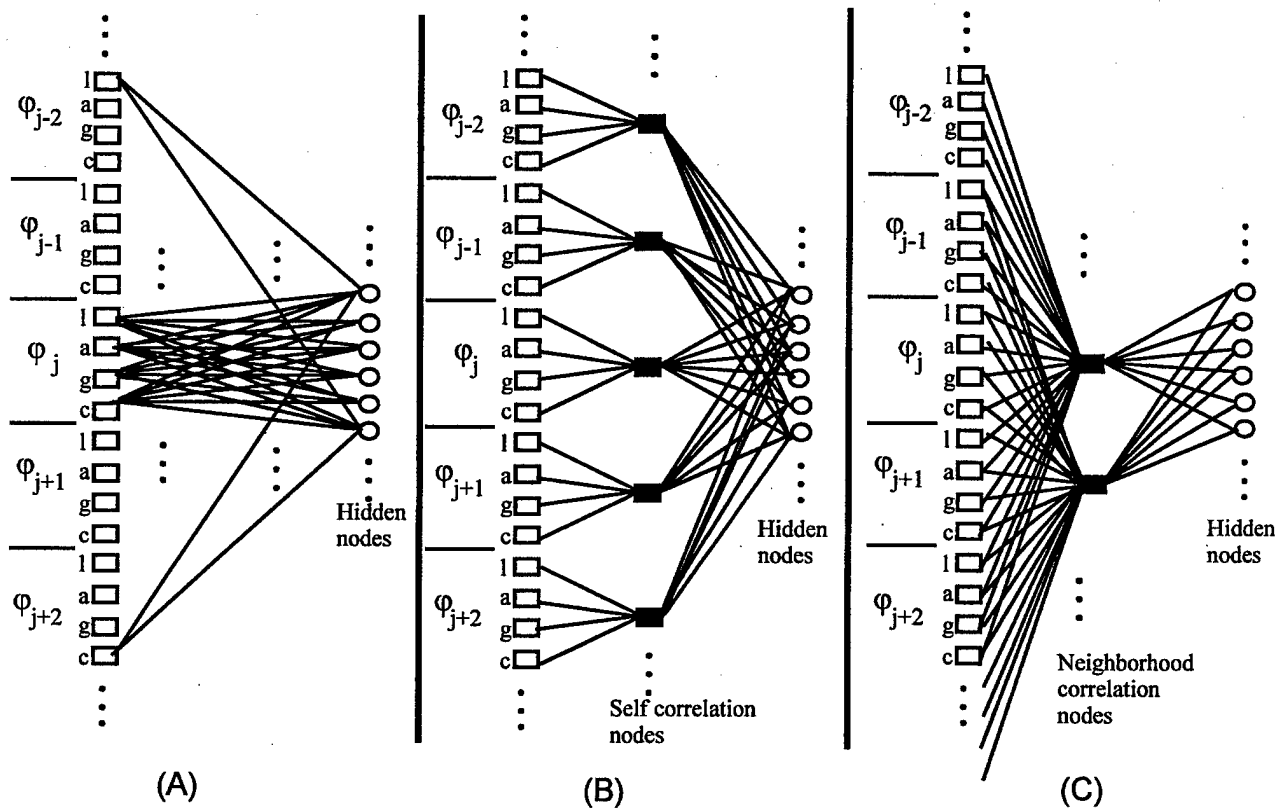


Figure 3. Three types of network paths connecting the input and the hidden layers:

(A) Full connection.

(B) A self correlation (SC) path; each node on the layer connects to a single set of the features (l,a,g,c) for the fan-in and fully connects to the hidden nodes for fan-out.

(C) A neighborhood correlation (NC) path; each node on the layer connects to five adjacent sets of the features for the fan-in and fully connects to the hidden nodes for fan-out.

Note that the fan-in nets emphasizing self correlation in (B) and neighborhood correlation in (C) represent convolution weights (i.e., the same type of sectors possess the same set of weighting factors).

3.4. Summary of feature arrangement methods and the MCPNN

We have described our approach on the feature extraction, the design of MCPNN, and its corresponding training method. Figure 4 shows a flow diagram of the proposed method. Since the MCPNN only alters the input data connection from the input to the first hidden layer, any learning algorithm can be applied within the neural network. For simplicity, we used the back propagation algorithm for both the conventional and proposed neural network systems in the following experiments.

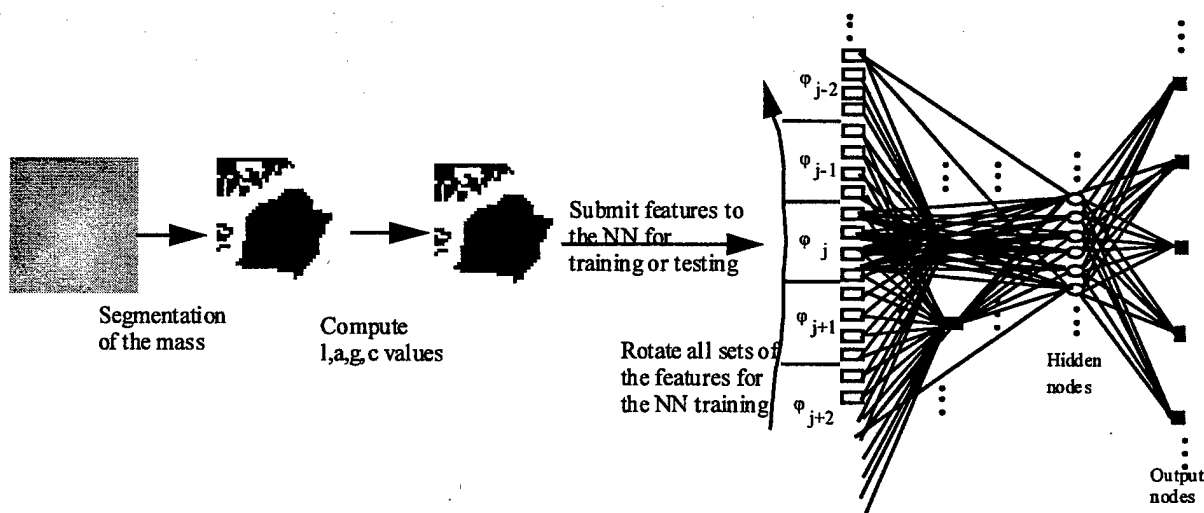


Figure 4. A schematic diagram, showing the MCPNN and sector features of masses, was used in the following study.

4. Experiments and Results

We selected 91 mammograms and digitized each mammogram with a computer format of $2048 \times 2500 \times 12$ bits (for an $8'' \times 11''$ area where each image pixel represents $100 \mu\text{m}$ square). No two mammograms were selected from the same patient film jacket. All the digitized mammograms were miniaturized to $512 \times 625 \times 12$ bits using 4×4 pixel averaging and were processed by the above methods to perform mass detection. Based on the corresponding biopsy reports, one experienced radiologist read all 91 mammograms and identified 75 areas containing masses. (Note that the reports recorded the malignancy of the biopsy specimens. The radiologist only used them as reference for the identification of masses.) Through the pre-process and the first step screen based on the circularity test, a total of 125 suspicious areas were extracted from the 91 digitized mammograms.

4.1. Experiment 1

We randomly selected 54 computer-segmented areas where 30 patches were matched with the radiologist's identification and 24 were not. This database was used to train two neural network systems: (1) a conventional 3-layer BP neural network (with 125 nodes in the hidden layer) and (2) the proposed MCPNN training method using the same neural network learning algorithm. The structure of the MCPNN was described earlier. However, we used one fully connected path, four SC paths, four 20° NC paths, four 30° NC paths, three 40° NC paths, and two 50° NC paths in the first step network connection for the MCPNN. All paths in the neural network have their hidden layers. Only one hidden layer per path was used. Both neural network systems were trained by the error back propagation algorithm by feeding the features from the input layer and registering the corresponding target value at the output side. Once the training of the neural networks was complete, we then used the remaining 71 computer segmented areas for the testing. None of the images and their corresponding patients in the testing set could be found in the training set. The neural network output values were fed into the LABROC program [21] for the performance evaluation. The results indicated that the areas (A_z) under the receiving operator characteristic (ROC) curves were 0.781 and 0.844 using the conventional BPNN and the MCPNN, respectively. The ROC curves of these two neural network training methods are shown in Figure 5(A). We also invited another senior mammographer to conduct an ROC observer study. The mammographer was asked to rate each patch using a numerical scale ranging 0-10 for its likelihood of being a mass. These 71 numbers were also fed into the LABROC program. The mammographer's performance in A_z on this set of test cases was 0.909. The corresponding ROC curve is also shown in Figure

5(A).

4.2. Experiment 2

We also conducted a leave-one-case-out experiment using the same database. In this experiment, we used those patches extracted from 90 mammograms for the training and used the patches (most of them are single) extracted from the remaining one mammogram as test objects. The procedure was repeated 91 times to allow every suspicious patch from each mammogram to be tested in the experiment. For each individual suspicious area, the computed features were identical to those used in Experiment 1. Again, both neural network systems were independently evaluated with the same procedure. The results indicated that the Az values were 0.799 and 0.887 using the conventional back propagation neural network and the MCPNN, respectively. Figure 5(B) shows the ROC curves of these two neural network systems using the leave-one-of-out procedure in the experiment.

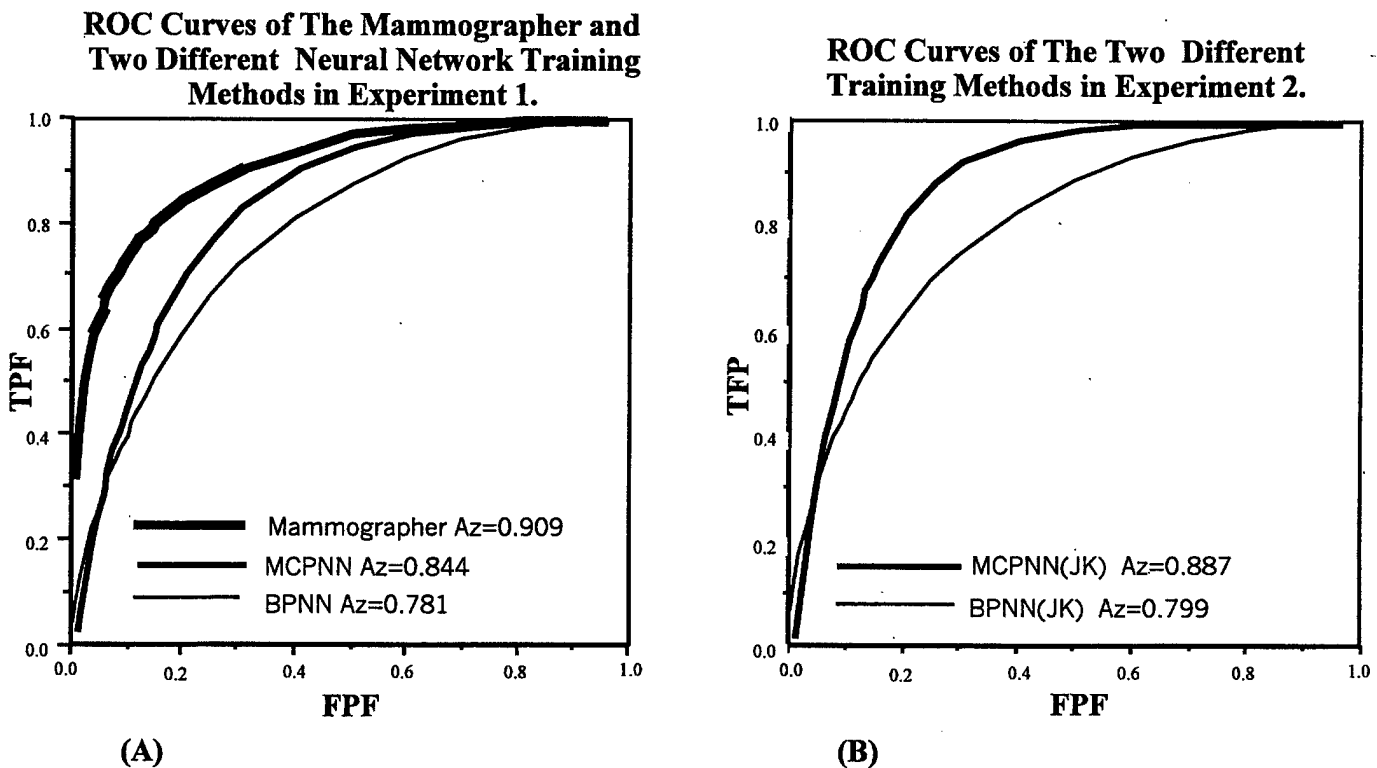


Figure 5. The ROC curves obtained from corresponding experiments.

- (A) The left figure shows that the performance of MCPNN training method is superior to that of the conventional input method. The highest curve is the ROC performance of the senior mammographer.
- (B) The right figure shows the ROC results with higher performance using the leave-one-case-out procedure as described in Experiment 2.

5. Conclusions and Discussion

Through this study, we found that the selected features are somewhat effective in the detection of masses. These features were "computationally translated" from the qualitative descriptors of BI-RAD. Another uniqueness of this study was on the test of our newly developed MCPNN training method. In Experiment 1, we found that the performances of both neural network systems were increased. This might be due to the increased number of cases (from 54 to 124) in the training set. In Experiment 2, the Az value was improved by 0.043 using the MCPNN training method that was higher than Az difference of 0.018 obtained by

the conventional training method. The results implied that the MCPNN learned more effectively than the conventional BP when the number of training cases was increased.

It is known in the field of artificial intelligence that the key factors in pattern recognition are: (1) effective methods in the extraction of features and (2) analytic methods (e.g., back propagation neural network) for the extracted features. In this study, we showed that the training method designed to guide the analyzer is also an important factor to a success of a pattern recognition task. Though this finding is not new, the research of developing training methods for various pattern recognition tasks has not established in the field of medical imaging. In this work, we demonstrated that organized features with proper network connection and task-oriented guidance would assist the neural network in performing the task.

As far as the research in recognition of masses is concerned, we believe that main concept of using sectors is an effective approach. **Note that any features arranged in the polar coordinate system can be trained by the MCPNN method.** Since the MCPNN only coordinates the input data, the internal neural network learning algorithm can be changed to other learning algorithms. A technique using the rubber band straightening transformation, independently developed by Sahnier et al. [22], for the detection of masses also employs a similar concept in extracting feature and/or texture in the polar coordinate. We believe that integration of effective feature and texture values computed at small sectors will be the research trend in mass detection. This paper focuses on neural network design and arrangement of features for effective pattern recognition of masses in medical imaging.

Acknowledgments

This work was supported by US Army Grant No. DAMD17-96-1-6254 (through a sub-grant from the University of Michigan, Ann Arbor). The content of this paper does not necessarily reflect the position or policy of the government. A part of the database, used in the study, was provided by Dr. Robert Shah of Brooke Army Medical Center. The LABROC program was written by Dr. C.E. Metz and his colleagues at the University of Chicago. The authors are also grateful to Ms. Susan Kirby for her editorial assistance.

References

1. L. Nystrom, L. E. Rutqvist, S. Wall, A. Lindgren, M. Lindqvist, S. Ryden, et. al., "Breast cancer screening with mammography: Overview of Swedish randomized trials," *Lancet*, vol. 341, pp. 973-978, 1993.
2. S. Shapiro, "Screening- Assessment of current studies," *Cancer*, vol. 74, pp. 231-238, 1994.
3. L. Tabar, G. Fagerberg, S. Duffy, N. E. Day, A. Gad, and O. Grontoft, "Update of the Swedish two-contry program of mammographic screening for breast cancer," *Radiology Clinics of North America: Breast Imaging - Current Status and Future Directions*, vol. 30, pp. 187-210, 1992.
4. S. M. Lai, X. Li, and W. F. Bischof, "On techniques for detecting circumscribed masses in mammograms," *IEEE Trans. Med. Imag.*, vol. 8, pp. 377, 1989.
5. D. Brzakovic, X. M. Luo, and P. Brzakovic, "An approach to automated detection of tumors in mammograms," *IEEE Trans. Med. Img.*, vol. 9, pp. 233, 1990.
6. F. F. Yin, M. L. Giger, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E. Metz, "Computerized detection and characterization system for use in mammographic screening programs," *Radiology*, vol. 177(P):245, 1990.
7. B. Sahiner, H. P. Chan, N. Petrick, D. Wei, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "Classification of mass and normal breast tissues: A convolution neural network classifier with spatial domain and texture images," *IEEE Trans. on Med. Imag.*, vol. 15, pp. 598-610, 1996.
8. D. Wei, H. P. Chan, M. A. Helvie, B. Sahiner, N. Petrick, D. D. Adler, and M. M. Goodsitt, "Classification of mass and normal breast tissue on digital mammograms: Multiresolution texture analysis," *Med. Phy.*, vol. 25, no. 4, pp. 516-526, 1998.
9. D. D. Adler, "Breast masses: differential diagnosis." In Feig SA, ed. *ARRS categorical course syllabus on breast imaging*. Reston, VA: American Roentgen Ray Society, p. 31, 1988.

10. M. J. Homer, "Imaging features and management of characteristically benign and probably benign breast lesions," *Radiol. Clin. North Am.*, vol. 25, p. 939, 1987.
11. M. Moskowitz, "Circumscribed lesions of the breast" In Moskowitz M, ed. *Diagnostic categorical course in breast imaging*. Oak Brook, Ill: Radiol. Soc. of North Am., p. 31, 1986.
12. E. A. Sickles, "The rule of multiplicity and the developing density sign." In: Feig SA, ed. *ARRS categorical course syllabus on breast imaging*. Reston, VA: Am. Roent. Ray Soc., p. 177, 1988.
13. H. Li, Y. Wang Y, K. J. R. Liu, S. C. Lo, and M. T. Freedman, "Statistical Model Supported Approach to Radiographic Mass Detection -- Part I: Improving Lesion Characterization by Morphological Filtering and Site Segmentation," to appear *IEEE Trans. on Med. Imag.*, 2001.
14. H. Li, Y. Wang Y, K. J. R. Liu, S. C. Lo, and M. T. Freedman, "Statistical Model Supported Approach to Radiographic Mass Detection -- Part II: Decision Making through Modular Neural Networks and Hierarchical Visual Explanation," to appear *IEEE Trans. on Med. Imag.*, 2001.
15. *Breast Imaging -- Reporting and Data System*, American College of Radiology, Reston, Virginia, 1993.
16. S. C. Lo, M. T. Freedman, J. S. Lin, and S. K. Mun, "Automatic lung nodule detection using profile matching and backpropagation neural network techniques," *J. of Dig. Imag.*, vol. 6(1), pp. 48-54, 1993.
17. S. C. Lo, H. P. Chan, J. S. Lin, H. Li, M. T. Freedman, and S. K. Mun, "Artificial convolution neural network for medical image pattern recognition," *Neural Networks*, 1995, Vol. 8, No. 7/8, pp. 1201-1214, 1995.
18. H. P. Chan, S. C. Lo, B. Sahiner, K. L. Lam, and M. A. Helvie, "Computer-aided diagnosis of mammographic microcalcifications: Pattern recognition with an artificial neural network," *Med. Phy.*, vol. 24, No. 10, pp. 1555-1567, 1995.
19. Y. Wu, K. Doi, M. L. Giger, & R. M. Nishikawa, "Computerized detection of clustered microcalcifications in digital mammograms: Applications of artificial neural networks," *Med. Phy.*, vol. 19, pp. 555-560, 1992.
20. Y. Wu, M. T. Freedman, S. C. Lo, R. A. Zuurbier, A. Hasegawa, and S. K. Mun, "Classification of microcalcifications in radiographs of pathological specimen for the diagnosis of breast cancer," *Acad. Radiol.*, pp. 199-204, 1995.
21. C. E. Metz, J. H. Shen, and B. A. Herman, "New methods for estimating a binormal ROC curve from continuously distributed test results," Presented at the 1990 Annual Meeting of the American Statistical Association, Anaheim, CA.
22. B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. M. Goodsitt, "Computerized characterization of masses on mammograms: The rubber band straightening transform and textures analysis," *Med. Phy.*, vol. 25(4), pp. 516-526, 1998